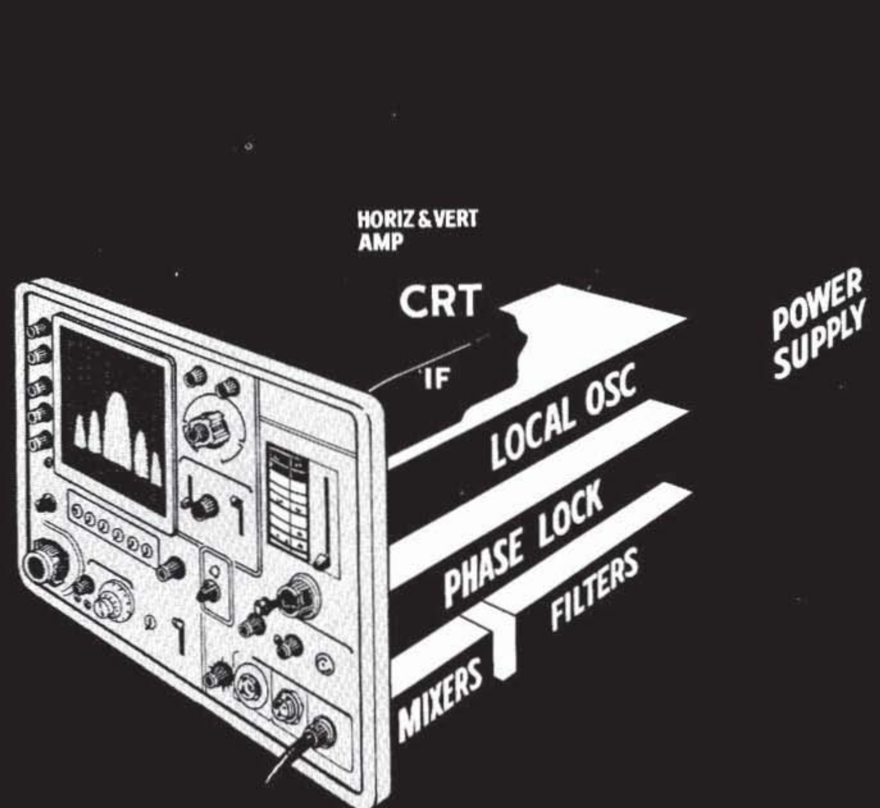
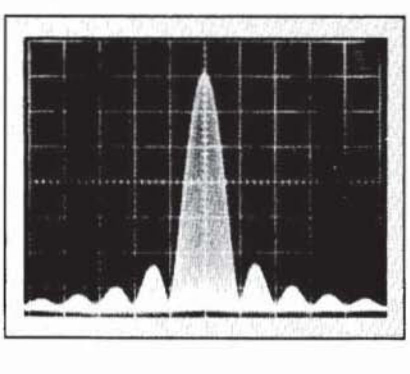
Circuit Concepts





Spectrum Analyzer Circuits

BOOKS IN THIS SERIES:

Circuit Concepts

Power Supply Circuits
062-0888-01

Oscilloscope Cathode-Ray Tubes
062-0852-01

Storage Cathode-Ray Tubes and Circuits
062-0861-01

Television Waveform Processing Circuits
062-0955-00

Digital Concepts
062-1030-00

Spectrum Analyzer Circuits
062-1055-00

Oscilloscope Trigger Circuits
062-1056-00

Sweep Generator Circuits
062-1098-00

Measurement Concepts

Information Display Concepts
062-1005-00

Semiconductor Devices
062-1009-00

Television System Measurements
062-1064-00

Spectrum Analyzer Measurements
062-1070-00

Engine Analysis
062-1074-00

Automated Testing Systems
062-1106-00

SPECTRUM ANALYZER CIRCUITS

BY

MORRIS ENGELSON

Significant Contributions
by
GOROON LONG
WILL MARSH

CIRCUIT CONCEPTS

FIRST EDITION, SECOND PRINTING, AUGUST 1969
062-1055-00

CONTENTS

1

BACKGROUND MATERIAL

Many spectrum analyzer circuits are based on principles with which most engineers may not be familiar. It is the purpose of this chapter to review some of this specialized background material. This review is not intended to be either complete or rigorous. Those desiring a more complete discussion are referred to the basic references and the bibliography.

TRANSMISSION LINES*

At low frequencies the basic circuit elements are lumped. At higher frequencies, however, where the size of circuit elements is comparable to a wavelength, lumped elements can not be used easily, if at all. This accounts for the extensive use of distributed circuits at higher frequencies.

A distributed circuit is a signal transmission path in which the basic circuit elements (resistance, inductance, capacitance) are distributed over the entire length of the transmission path. A transmission line constitutes a distributed circuit.

---

*This section pertains to uniform transmission lines; tapered transmission lines will not be considered.

electrical characteristics

Transmission lines can be constructed in many different ways. These will have different physical characteristics and certain relatively minor electrical differences. For example, the unshielded two-wire construction is susceptible to pickup and radiation, whereas the shielded coaxial structure is not. All transmission lines, however, have certain major electrical characteristics as follows: The transmission line is considered as consisting of an infinite number of elements composed of infinitesimal shunt capacitance ($C\Delta\ell$) and conductance ($G\Delta\ell$), and series resistance ($R\Delta\ell$) and inductance ($L\Delta\ell$). For most applications one can make the simplifying assumption that the transmission line is lossless by neglecting the effects of shunt conductance and series resistance.

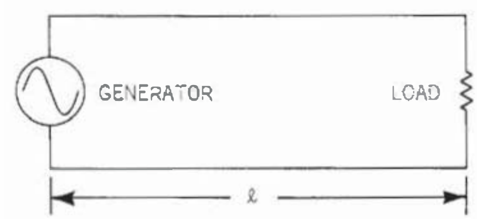


Fig. 1-1. Transmission Line.

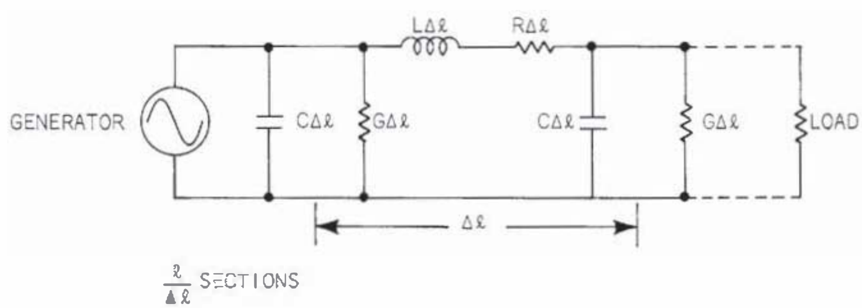


Fig. 1-2. Transmission-line equivalent circuit.

QUANTITY	SYMBOL	GENERAL EXPRESSION	IDEAL LOSSLESS LINE
CHARACTERISTIC IMPEDANCE	Z_0	$\sqrt{\frac{R + j\omega L}{G + j\omega C}}$	$\sqrt{\frac{L}{C}}$
PROPAGATION CONSTANT	$\gamma = \alpha + j\beta$	$\sqrt{(R + j\omega L)(G + j\omega C)}$	$j\omega\sqrt{LC}$
PHASE CONSTANT	β	(IMAGINARY PART OF γ)	$\omega\sqrt{LC} = \frac{2\pi}{\lambda}$
ATTENUATING CONSTANT	α	(REAL PART OF γ)	0
INPUT IMPEDANCE	Z_i	$Z_0 \frac{Z_r + Z_0 \tanh \gamma \ell}{Z_0 + Z_r \tanh \gamma \ell}$	$Z_0 \frac{Z_r + jZ_0 \tan \beta \ell}{Z_0 + jZ_r \tan \beta \ell}$
IMPEDANCE: LINE SHORTED		$Z_0 \tanh \gamma \ell$	$jZ_0 \tan \beta \ell$
LINE OPEN		$Z_0 \coth \gamma \ell$	$-jZ_0 \cot \beta \ell$
LINE = ODD NUMBER QUARTER WAVES		$Z_0 \frac{Z_r + Z_0 \coth \alpha \ell}{Z_0 + Z_r \coth \alpha \ell}$	$\frac{Z_0^2}{Z_r}$
LINE = EVEN NUMBER QUARTER WAVES		$Z_0 \frac{Z_r + Z_0 \tanh \alpha \ell}{Z_0 + Z_r \tanh \alpha \ell}$	Z_r
VOLTAGE REFLECTION COEFFICIENT	Γ	$\frac{Z_r - Z_0}{Z_r + Z_0}$	$\frac{Z_r - Z_0}{Z_r + Z_0}$
VOLTAGE STANDING WAVE RATIO	VSWR	$\frac{1 + \|\Gamma\|}{1 - \|\Gamma\|}$	$\frac{1 + \|\Gamma\|}{1 - \|\Gamma\|}$

R,L,G,C = DISTRIBUTED RESISTANCE, INDUCTANCE, CONDUCTANCE, CAPACITANCE PER UNIT LENGTH.

ℓ = DISTANCE ALONG LINE, MEASURED FROM LOAD END.

Z_r = LOAD AT RECEIVING END OF LINE, Z_r SHORTED LINE = 0, Z_r OPEN LINE = ∞.

$|\Gamma|$ = MAGNITUDE OF VOLTAGE REFLECTION COEFFICIENT, Γ.

Fig. 1-3. Basic transmission-line relationships.

When a transmission line is terminated in its characteristic impedance ($Z_r = Z_0$), the combination behaves as an infinitely long line. The energy delivered by the generator propagates down the transmission line and is absorbed by the load without reflections. When the load is other than the characteristic impedance, Z_0, then some of the incident energy is reflected back down the line toward the generator. This gives rise to peaks and nulls of voltage (and current) along the line known as standing waves.

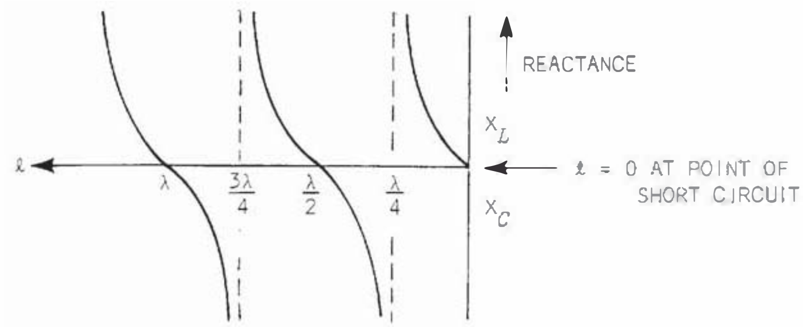


Fig. 1-4. Impedance of short-circuited lossless transmission line.

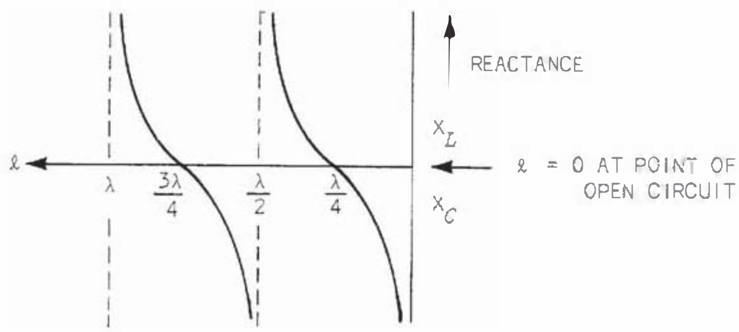


Fig. 1-5. Impedance of open-circuited lossless transmission line.

standing waves

The phenomenon of standing waves is easily understood in connection with the impedance of the transmission line. Figs. 1-4 and 1-5 show the variation of impedance as a function of line length for shorted and open-circuited lines respectively. It is obvious that the voltage has to go to zero at a zero-impedance point and toward infinity at a point of infinite impedance. In practice real transmission lines are not entirely lossless, so that the voltage along the line does not actually go to zero or become infinite. Note that the line impedance and, therefore, the standing-wave pattern repeats at half-wavelength $(\frac{\lambda}{2})$ intervals. This characteristic is independent of the impedance of the load.

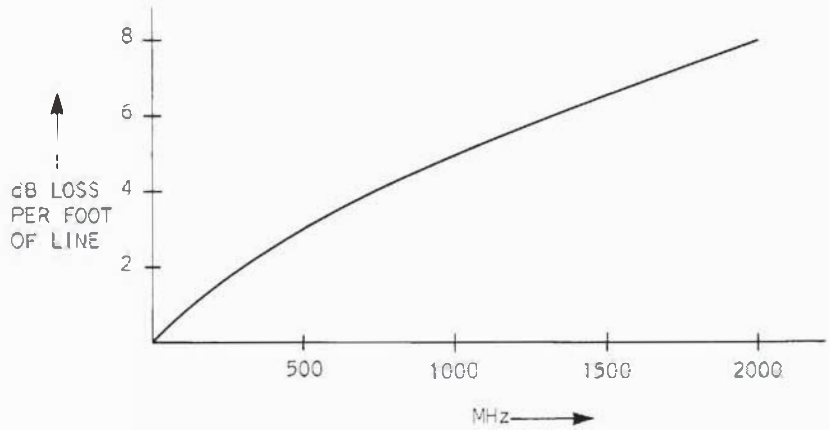


Fig. 1-6. Lossy transmission-line characteristic.

In most applications it is desirable that the transmission line have as little loss as possible. There are, however, applications where lossy transmission lines work best. Lossy transmission lines are manufactured by many companies including Tektronix, Inc. The Tektronix lossy transmission line is of coaxial construction where the center conductor is made of steel piano wire. The steel center conductor has the useful property of increasing insertion loss with increasing frequency, as shown in Fig. 1-6. The power loss increases approximately as the square root of frequency. Such transmission lines are useful in applications where the energy of interest is at relatively low frequencies or the loss can be tolerated, and a good high-frequency match is important. A length of lossy transmission line will provide a good impedance match, low VSWR, even when the load is much different from the characteristic impedance, Z_0. This is because the reflected wave is attenuated by the lossy line. Thus, the voltage reflection coefficient Γ decreases as we go further from the mismatched load: $|\Gamma_\ell| = |\Gamma_0|\varepsilon^{-2\alpha\ell}$. Where: $|\Gamma|$ is the absolute value of the voltage reflection coefficient at any point along the line, and $|\Gamma_0|$ is absolute value of voltage reflection coefficient at zero distance from the load.

lossy transmission lines

Γ, voltage reflection coefficient

For example, $\Gamma_0 = -1$ for a short circuit; one foot of Tektronix lossy cable will reduce Γ to about 0.16 at 2 GHz and VSWR will be reduced from an infinite ratio to 1.4.

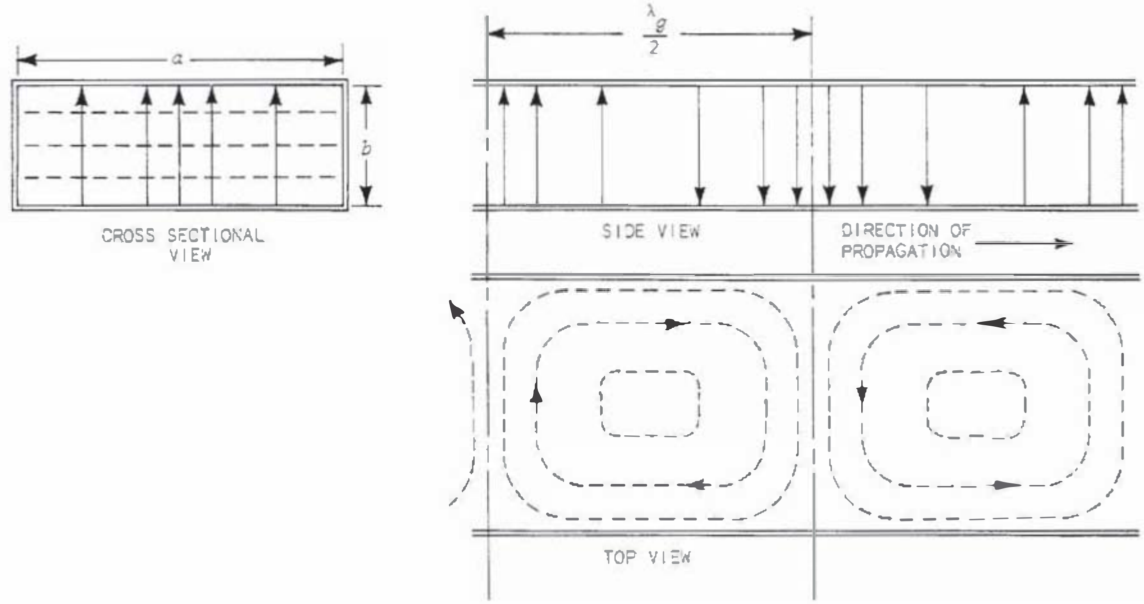


Fig. 1-7. $TE_{1,0}$ fields in a rectangular waveguide. Solid lines -- electric field; dotted lines -- magnetic field.

WAVEGUIDES

In principle, a waveguide is any structure which serves the function of guiding electromagnetic energy from one place to another. Transmission lines could, therefore, be considered as a specific type of waveguide. In practice the word waveguide usually refers to a hollow metal pipe, and that is the definition we shall use here.

The advantages of waveguides over transmission lines are: 1) Higher power handling capability, 2) Lower losses, 3) Easier and more rugged construction, 4) It is almost impossible to build transmission lines that do not propagate in undesirable modes at higher frequencies, hence the predominance of waveguides above 12.4 GHz.

propagation modes

All transmission systems, both waveguide and transmission line, have an infinite number of *modes* by which energy can be propagated. Each of these modes has a *cutoff* frequency below which energy cannot be propagated. Modes are specified as TE (Transverse Electric Field), where no component of *electric* field is in the direction of propagation; TM (Transverse Magnetic Field), where no component of *magnetic* field is in the direction of propagation; and TEM (Transverse Electromagnetic), where *neither* the electric nor magnetic field is in the direction of propagation.

cutoff frequency and mode

Transmission systems behave as high-pass filters where the cutoff frequency is a function of the mode of propagation. The higher the order of the mode the higher the cutoff frequency. In transmission lines the dominant mode is TEM which has no cutoff, implying cutoff at DC, so that transmission lines have essentially no low-frequency limit.

Most rectangular waveguides operate in the $TE_{1,0}$ mode. The field distribution for this mode is shown in Fig. 1-7.

The cutoff frequency for each mode can be calculated in terms of the waveguide dimensions a and b. The relationship is:

$$\lambda_c = \frac{v_c}{f_c} = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}}$$

cutoff frequency and dimensions

where λ_c and f_c are the cutoff wavelength and cutoff frequency respectively for the $TE_{m,n}$ or $TM_{m,n}$ mode, and v_c is the velocity of light (3×10^8 m/s or 11.8×10^9 in/s).

For the dominant $TE_{1,0}$ mode, $\lambda_c = 2a$. For example, WR-62 waveguide is recommended for the 12.4 to 18-GHz frequency range. The dimensions of this waveguide are 0.622 by 0.311 inches. The cutoff frequency for the $TE_{1,0}$ mode computed from $f_c = \frac{v_c}{2a}$ is $\frac{11.8 \times 10^9}{1.244}$ or 9.486 GHz. Note that the recommended lower operating frequency is well above the cutoff frequency. The recommended upper operating frequency is determined on the basis that no higher-order modes propagate. In practice, waveguide operating-wavelength ranges are usually between 60 and 95% of the cutoff.

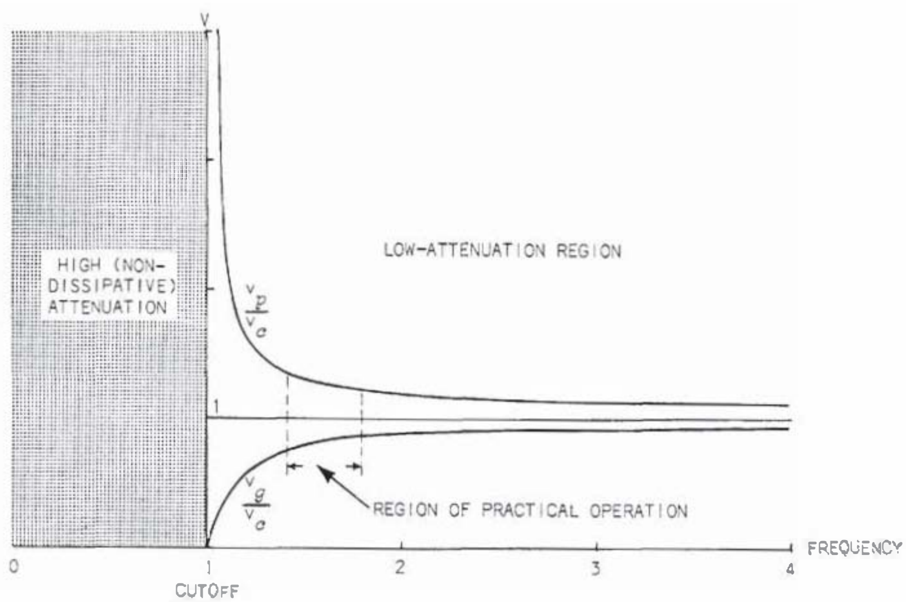


Fig. 1-8. Normalized phase velocity and group velocity versus normalized frequency.

propagation velocity

Unlike a transmission line, the propagation velocity down a waveguide is not constant, but varies with frequency. Two velocities are involved. One is the phase velocity (v_p). This is an apparent or fictitious velocity which pertains to the phase front of the propagating electromagnetic wave. The other is the group velocity (v_g). This is the rate at which the electromagnetic wave actually transfers power. Fig. 1-8 is a plot of these velocities as a function of frequency.

phase velocity

group velocity

Unlike the transmission line, one normally does not consider a waveguide as having a distinct characteristic impedance, since distinct voltages and currents do not exist. Instead, one considers the power propagating down the guide. The concepts of reflection coefficient, VSWR and optimum power match are, however, applicable; since they depend on ratios rather than absolute levels of voltage or current.

CAVITY RESONATORS

Fig. 1-9 shows a typical low-frequency resonant circuit. The resonant frequency is $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and the Q-factor which is a function of circuit losses is $Q = \frac{R}{2\pi f_0 L}$ with a half-power (3-dB) bandwidth of $\Delta f = \frac{f_0}{Q}$.

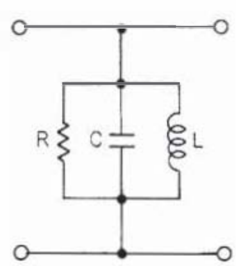


Fig. 1-9. Parallel resonant circuit.

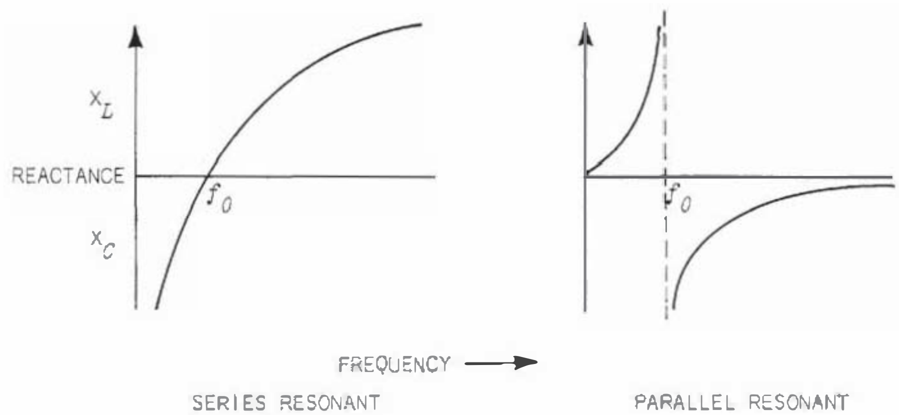


Fig. 1-10. Reactance vs frequency resonant circuits.

resonant circuits and cavities

As previously indicated, lumped-parameter circuits are not practical at higher frequencies. This makes it necessary to build distributed resonators. Referring to Figs. 1-4 and 1-5 observe by comparison to the reactance characteristics of lumped resonant circuits (Fig. 1-10), that the open- and the short-circuited transmission lines are distributed resonant circuits.

series and parallel resonance

However, where the lumped-parameter circuit has a single resonant frequency, the transmission-line resonator has an infinite number of resonant frequencies. Thus a short-circuited transmission line has parallel resonance characteristics every odd quarter-wavelength ($\frac{\lambda}{4}$, $\frac{3\lambda}{4}$, etc.) and series resonance characteristics at even quarter-wavelengths ($\frac{2\lambda}{4}$, $\frac{4\lambda}{4}$, etc.). These resonances are referred to as modes, e.g., $\frac{3\lambda}{4}$ mode.

waveguide cavity

Waveguide cavities are used in applications where transmission lines are not suitable (e.g., frequency too high or too much loss). These consist, essentially, of a short-circuited piece of waveguide forming a closed box. Just as for a waveguide, a waveguide cavity has an infinite number of TM and TE modes. These are characterized by three numbers, the

first two referring to the waveguide mode and the third indicating the length of the cavity in half-guide wavelengths. Thus a $TM_{2,1,2}$ mode in a rectangular cavity means a $TM_{2,1}$ waveguide mode, and a cavity two half-guide-wavelengths (or λ_g) long. A waveguide cavity can have any shape, though most practical designs utilize either rectangular or cylindrical cavities.

selecting modes

Since a cavity will normally support many modes, steps must be taken to reinforce the desired mode and eliminate the others. This takes the form of slots and absorptive material so placed as to prevent current flow and attenuate undesired modes, and selective coupling so as to couple to the desired mode only. Some of the different coupling methods are illustrated in Fig. 1-11.

cavity Q

The concepts of resonant frequency, Q, and bandwidth are applicable to waveguide cavities. The Q-factor, which is proportional to the ratio of energy stored per cycle to power lost, is dependent on the frequency, conductivity of the cavity material and on the ratio of V/A where V = cavity volume and A = cavity area.

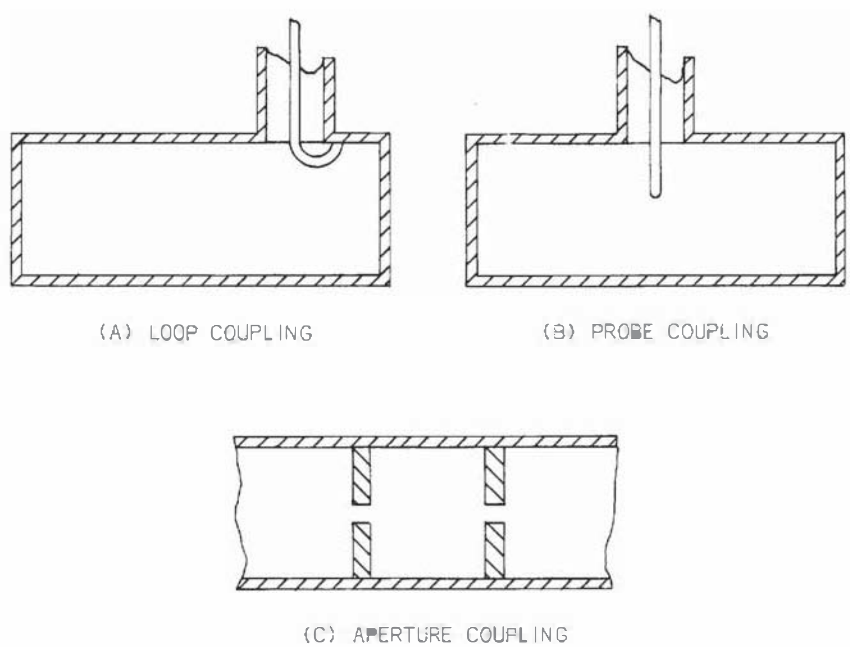


Fig. 1-11. Coupling techniques.

Thus, maximum Q is obtained for a cube, and for a cylinder with length equal diameter, since these configurations maximize V/A. The Q of a cavity is affected (lowered) by the coupling structure. It is therefore necessary to differentiate between the unloaded Q (Q_U) and the loaded Q (Q_L). These are related by $\frac{1}{Q_L} = \frac{1}{Q_U} + \frac{1}{Q_E}$, where Q_E is the Q of the external load. As for other circuits the 3-dB bandwidth is determined by the Q_L, namely $\Delta f = \frac{f_0}{Q_L}$.

unloaded, loaded Q

CLOSED-LOOP CONTROL CIRCUITS

Feedback is a well understood principle as applied to amplifiers and oscillators.

feedback-control systems

The application of inverse (cancelling, negative, etc.) feedback in amplifiers has the effect of increasing bandwidth, gain stability, and tolerance to parameter variations.

The application of positive feedback has the opposite effect, and in general, when the resulting loop gain is greater than one, will result in oscillation.

Feedback-control theory is extremely complex and requires an understanding of transform mathematics for design and analysis of all but the most basic systems. It is our intention to give here only a qualitative understanding of the more complex feedback-control systems encountered in spectrum analyzers.

stability improvement

The function of most feedback-control systems in electronic instruments is: 1) To linearize a control function, 2) To improve both long- and short-term stability, 3) To reduce the effect of parameter variations which would otherwise have deleterious effects.

control-system analysis

feedback voltage

control voltage

First, let us assume that the overall loop gain of the system is very large in comparison to the output/input ratio. Only a small (indeed, infinitesimal) difference between the control-voltage value and the feedback-voltage value is required to effect whatever control is required. The required amount of output is achieved from the controlled device by a vanishingly small difference between control voltage and feedback voltage, or finally and most important: since the feedback voltage *is essentially equal to the control voltage* and the output is related to the feedback voltage by the sensing device, then -- *the relationship between the output and input is essentially determined by the sensing device.*

high loop gain

This is an extremely important conclusion. So long as the loop gain is sufficiently high, the characteristics of the control system are essentially determined by the sensing device, and are essentially independent of the characteristics of the device to be controlled.

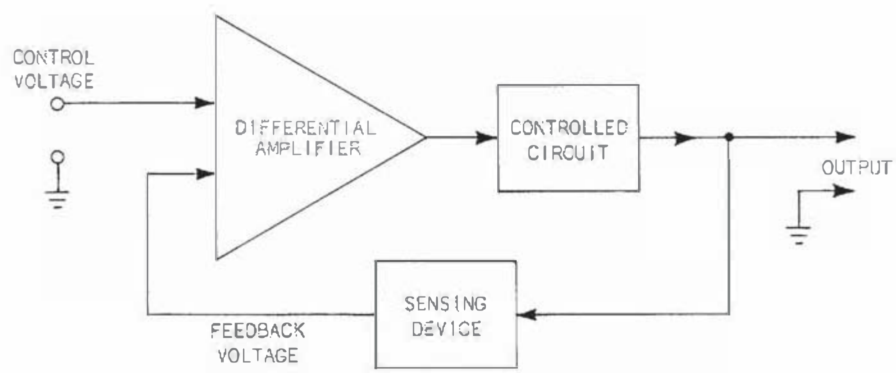


Fig. 1-12. Block diagram of feedback-control system.

A few ways in which these properties may be exploited are illustrated below.

Imagine that the controlled device is a voltage-controlled oscillator, varactor-tuned for example, and that the sensing device is a crystal discriminator.

varactor, discriminator

This oscillator is a relatively unstable and nonlinear device. It is unstable because Q is likely to be relatively low, leading to frequency shifts by transistor parameters, and because the varactor has an appreciable shift of capacitance with temperature. It is nonlinear because most of the varactor-capacitance change occurs at low voltages with very little C change at high voltages.

The discriminator, on the other hand, can be an extremely linear and stable device, perhaps two orders of magnitude more stable and linear than the oscillator.

gain in stability

The controlled output, in this case the relationship between the output frequency and input voltage, can be almost two orders of magnitude more stable and linear than it would have been in a nonfeedback control system.

This illustration is, indeed, a common application of feedback in swept-frequency oscillators in spectrum analyzers.

phase-lock control

A slightly more complex application of feedback occurs in phase-locked oscillators. Suppose that the reference strobe (Fig. 1-13) consists of very short pulses in relation to the voltage-controlled oscillator period and that the difference voltage between the two inputs to the differential amplifier needed to exercise the required control on the oscillator is small in relation to the phase-control signal.

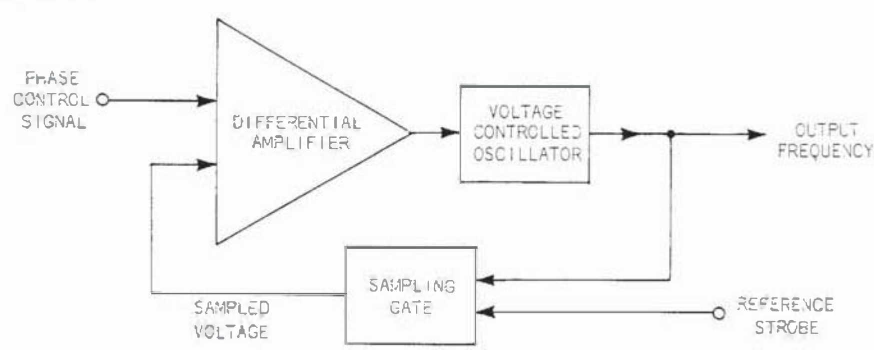


Fig. 1-13. Phase-lock oscillator feedback control system.

oscillator synchronizes with strobe pulses

Then the relationship between the timing of the strobe pulses and a point on the oscillator waveform must be such that the output of the sampling gate is equal, essentially, to the phase-control voltage. This can happen only if the oscillator output is stationary in time with respect to the strobe pulses, or in other words, locked in phase to the strobe pulses.

These are rudimentary examples of the use and functioning of feedback-control systems in spectrum analyzers. A lot has been intentionally omitted, especially the effects of response rate of system components and noise. For those who wish to dig deeper, see the references cited.

SPECTRUM ANALYZER SYSTEMS

power distribution by frequency

According to the Tektronix definitions of terms, a spectrum analyzer is a device which displays a graph of relative power distribution as a function of frequency, typically on a cathode-ray tube or chart recorder. This definition encompasses many different devices. A properly programmed computer, a bank of filters, a swept or tuned filter, or a superheterodyne signal-translation system would all be considered spectrum analyzers as judged by the above definition. This book will concentrate on the superheterodyne signal-translation method of spectrum analysis. This is what the words *spectrum analysis* will be intended to convey unless otherwise indicated.

The heart of the superheterodyne signal-translation process is the mixer, sweeping oscillator, and narrowband filter system, as shown in Fig. 1-14. This block is positioned in different places along the signal processing chain, as dictated by the instrument performance requirements and the economics of the situation. When the sweeping oscillator is the first in the superheterodyne chain, the system is called

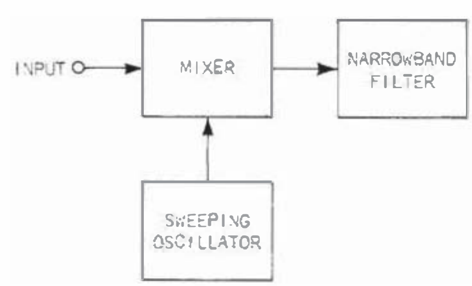


Fig. 1-14. Basic spectrum-analyzer block.

swept front end, IF

swept front end. Similarly, when the sweeping oscillator is positioned further down the signal processing chain, the system is called *swept IF*. The theory of operation is essentially the same for both systems, the difference being mainly in the performance parameters.

frequency-time analogue

Basically the superheterodyne signal-sweeping spectrum analyzer performs a spectrum analysis by means of a signal being translated in frequency past a stationary filter. For example, a frequency-translating CW signal sweeping past a stationary filter will produce an output pulse. The pulse width is determined by the total time that the sweeping-signal frequency dwells within the filter bandwidth. This provides an analog representation of signal strength versus frequency, where signal strength is represented by pulse height, and frequency is represented by pulse position on a calibrated time scale. The pulse width, which is a function of sweep time and filter bandwidth, determines the resolution of the system; i.e., how close in frequency can two signals be before we are unable to tell them apart.

A practical spectrum analyzer contains many circuits besides those shown in Fig. 1-14. These would include such things as variable-bandwidth filters, crystal oscillators, wideband amplifiers, linear and logarithmic detectors, and others. The operation and function of these circuits is discussed in detail in the chapters that follow.

DEFINITIONS OF SPECTRUM ANALYZER TERMS

Spectrum Analyzer -- A device which displays a graph of relative power distribution as a function of frequency, typically on a cathode-ray tube or chart recorder.

A. Real Time. A spectrum analyzer that performs a continuous analysis of the incoming signal with the time sequence of events preserved between input and output.

B. Nonreal Time. A spectrum analyzer that performs an analysis of a repetitive event by a sampling process.

1. Swept front-end spectrum analyzer. A superheterodyne spectrum analyzer in which the first local oscillator is swept.

2. Swept intermediate-frequency spectrum analyzer. A superheterodyne spectrum analyzer in which a local oscillator other than the first is swept.

Center Frequency (radio frequency or intermediate frequency) -- That frequency which corresponds to the center of the reference coordinate (in units of Hz).

Center Frequency Range (radio frequency) -- That range of frequency that can be displayed at the center of the reference coordinate. When referred to a control (e.g., Intermediate Frequency Center Frequency Range), the term indicates the amount of frequency change available with the control (in units of Hz).

Deflection Factor -- The ratio of the input signal amplitude to the resultant displacement of the indicating spot (e.g., RMS V/div).

Dispersion (sweep width) -- The frequency sweep excursion over the frequency axis of the display. Can be expressed as frequency/full frequency axis or frequency/div in a linear display.

Drift (frequency drift) (stability) -- Long-term frequency changes or instabilities caused by frequency changes in the spectrum-analyzer local oscillators. Drift limits the time interval that a spectrum analyzer can be used without retuning or resetting the front panel controls (units may be Hz/s, Hz°C, etc.).

Display Flatness -- Uniformity of amplitude response over the rated maximum dispersion (usually in units of dB).

Dynamic Range (on screen) -- The maximum ratio of signal amplitudes that can be simultaneously observed within the graticule (usually in units of dB).

Dynamic Range, Maximum Useful -- The ratio between the maximum input power and the spectrum analyzer sensitivity (usually in units of dB).

Frequency Band -- A range of frequencies that can be covered without switching (in units of Hz).

Frequency Scale -- The range of frequencies that can be read on one line of the frequency indicating dial (in units of Hz).

Incidental Frequency Modulation (residual frequency modulation) -- Short-term frequency jitter or undesired frequency deviation caused by instabilities in the spectrum-analyzer local oscillators. Incidental frequency modulation limits the usable resolution and dispersion (in units of Hz).

Incremental Linearity -- A term used to describe local aberrations seen as nonlinearities for narrow dispersions.

Linearity (dispersion linearity) -- Measure of the comparison of frequency across the dispersion to a straight-line frequency change. Measured by displaying a quantity of equally spaced (in frequency) frequency markers across the dispersion and observing the positional deviation of the markers from an idealized sweep as measured against a linear graticule. Linearity is within $\frac{\Delta W}{W} \times 100\%$ where ΔW is maximum positional deviation and W is the full graticule width.

Maximum Input Power -- The upper level of input power that the spectrum analyzer can accommodate without degradation in performance (e.g., spurious responses and signal compression)(usually in units of dBm).

Maximum Sensitivity --

A. Signal equals noise. That input signal level (usually in dBm) which results in a display where the signal level above the residual noise is equal to the residual noise level above the baseline; expressed as: signal + noise = twice noise.

B. Minimum discernible signal. That input signal level (usually in dBm) which results in a display where the signal is just distinguishable from the noise.

Minimum Usable Dispersion -- The narrowest dispersion obtainable for meaningful analysis. Defined as ten times the incidental frequency modulation when limited by "incidental frequency modulation" (in units of Hz).

Optimum Resolution -- The best resolution obtainable for a given dispersion and a given sweep time (in units of Hz), theoretically:

$$\text{Optimum Resolution} = \sqrt{\frac{\text{dispersion (in Hz)}}{\text{sweep time (in seconds)}}}$$

Optimum Resolution (bandwidth) -- The bandwidth at which best resolution is obtained for a given dispersion and a given sweep time (in units of Hz):

$$\text{Optimum Resolution (bandwidth)} = 0.66\sqrt{\frac{\text{dispersion}}{\text{sweep time}}}$$

Resolution -- The ability of the spectrum analyzer to display adjacent signal frequencies discretely. The measure of resolution is the frequency separation of two equal amplitude signals, the displays of which merge at the 3-dB down points (in units of Hz).

The resolution of a given display depends on three factors: sweep time, dispersion and the bandwidth of the most selective amplifier. The 6-dB bandwidth of the most selective amplifier (when Gaussian) is called resolution bandwidth and is the narrowest bandwidth that can be displayed as dispersion and sweep time are varied. At very long sweep times, resolution and resolution bandwidth are synonymous.

Resolution (bandwidth) -- Refer to resolution.

Safe Power Level -- The upper level of input power that the spectrum analyzer can accommodate without physical damage (usually in units of dBm).

Scanning Velocity -- Product of dispersion and sweep repetition rate (in units of Hz/unit time).

Sensitivity -- Rating factor of spectrum analyzers ability to display weak signals.

Skirt Selectivity -- A measure of the resolution capability of the spectrum analyzer when displaying signals of unequal amplitude. A unit of measure would be the bandwidth at some level below the 6-dB down points, (e.g., 10, 20, 40-dB down) (in units of dB).

Spurious Response (spurii, spur) -- A characteristic of a spectrum analyzer wherein displays appear which do not conform to the calibration of the radio frequency dial. Spurii and spur are the colloquialisms used to mean spurious responses (plural) and spurious response (singular) respectively. Spurious responses are of the following type:

A. Intermediate frequency feedthrough. Wherein signals within the intermediate frequency passband of the spectrum analyzer reach the intermediate frequency amplifier and produce displays on the cathode-ray tube that are not tunable with the radio-frequency center-frequency controls. These signals do not enter into a conversion process in the first mixer and are not affected by the first local-oscillator frequency.

B. Image responses. When the input signal is above or below the local-oscillator frequency by the intermediate frequency, the superheterodyne process results in two major responses separated from each other by twice the intermediate frequency. The spectrum analyzer is usually calibrated for only one of these responses. The other is called the image.

C. Harmonic conversion. The spectrum analyzer will respond to signals that mix with harmonics of the local oscillator and produce the intermediate frequency. Most spectrum analyzers have dials calibrated for some of these higher-order conversions. The uncalibrated conversions are spurious responses.

D. Intermodulation. In the case of more than one input signal, the myriad of combinations of the sums and differences of these signals between themselves and their multiples creates extraneous responses known as intermodulation. The most harmful intermodulation is third order, caused by the second harmonic of one signal combining with the fundamental of another.

E. Video detection. The first mixer will act as a video detector if sufficient input signal is applied. A narrow pulse may have sufficient energy at the intermediate frequency to show up as intermediate frequency feedthrough.

F. Internal. A display shown on the cathode-ray tube caused by a source or sources within the spectrum analyzer itself and with no external input signal. Zero frequency feedthrough is an example of such a spurious response.

G. Anomalous IF responses. The filter characteristic of the resolution-determining amplifier may exhibit extraneous passbands. This results in extraneous spectrum-analyzer responses when a signal is being analyzed.

Sweep Repetition Rate -- The number of sweep excursions per unit of time. Sometimes approximated as the inverse of sweep time for a free-running sweep.

Sweep Time -- The time required for the spot in the reference coordinate (frequency in spectrum analyzers) to move across the full graticule width (can be expressed as TIME/DIV in a linear system).

Zero Frequency Feedthrough (zero pip) -- The response of a spectrum analyzer which appears when frequency of the first local oscillator is equal to the intermediate frequency. This corresponds to zero input frequency and is sometimes deliberately not suppressed so as to act as a zero frequency marker.

REFERENCES; see page 171.

Transmission Lines -- B-4, B-5, B-13 Chapter 20, C-4 pp18-69
Waveguides -- B-6, B-13 Chapter 21
Cavities -- B-4 Chapter 7, B-13 Chapter 21
Feedback Control Systems -- C-1, C-2

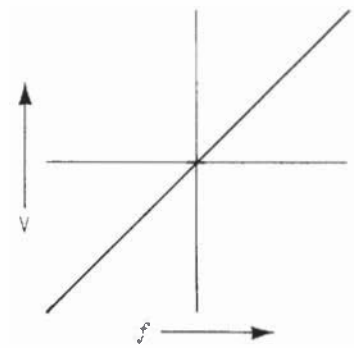


Fig. 2-1. Discriminator-transfer characteristic.

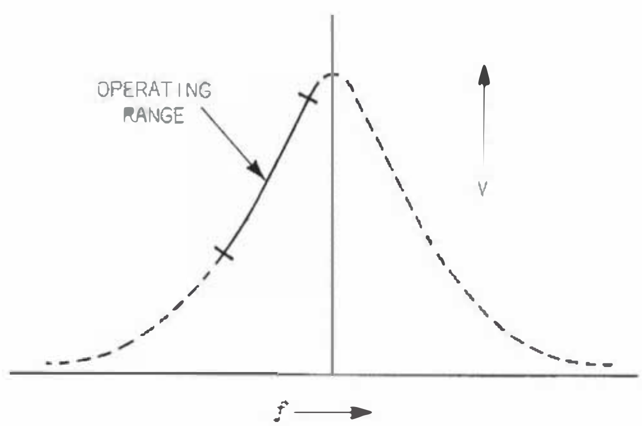


Fig. 2-2. Tuned circuit as a discriminator.

2

COMPONENTS AND SUBASSEMBLIES

DISCRIMINATORS

discriminator
demodulator

A discriminator is a device that converts a constant-amplitude frequency-varying signal into an output whose amplitude is proportional to the frequency deviation. The discriminator is best known as a demodulator for FM radio receivers. Fig. 2-1 shows an ideal discriminator.

Any tuned circuit with a detector can be used as a primitive type of discriminator, as shown in Fig. 2-2. However, single tuned circuits do not make good discriminators. Problems include, among others, poor linearity and difficulty in broad banding. As a result practical discriminators usually contain several tuned circuits plus various peripheral sub-circuits such as, center frequency and linearity controls, and detectors and low-pass filters for carrier elimination.

Fig. 2-3 shows a practical discriminator circuit using two parallel tuned LC circuits. The frequency characteristics of this discriminator are demonstrated in Fig. 2-4.

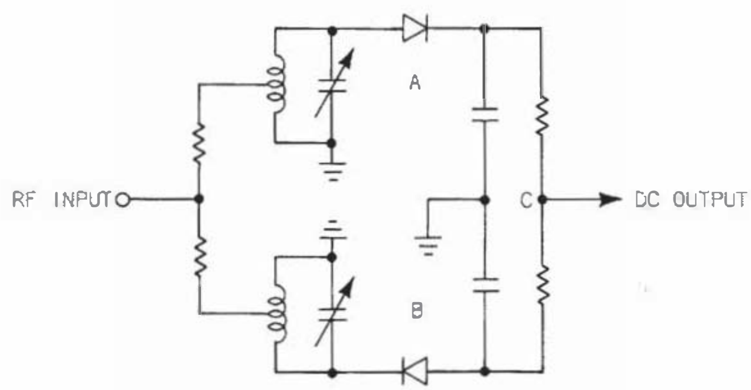


Fig. 2-3. LC discriminator.

Fig. 2-4A and 2-4B are the resonance curves of the two tank circuits which are tuned slightly above and below the discriminator band edges respectively. The detected and filtered outputs are combined at point C to yield the discriminator response as shown. Note that one of the detector diodes is inverted to present the proper polarity output to the summing point.

design criteria

Basic design criteria are: Linearity of input-output relationships, balance of symmetry for positive- and negative-going outputs, and sensitivity in volts out per ΔHz at a specified input level. Improved performance in these areas can be obtained by the addition of adjustments. Such adjustments include input and output potentiometers for balance control. These would be used to compensate for differences between the two channels (A and B) caused by mismatch between the diodes and unequal Q's in the tank circuits. The Q's of the tank circuits could be adjusted by paralleling with either fixed or variable resistors. This affects the balance, the linearity and the sensitivity or gain (volts out per Hz deviation in, at a specified constant input level). In addition to adjustments the basic circuit comes in several configurations. For example, the two channels may be summed in an amplifier rather than as described previously. Under such conditions the necessary polarity inversion of Channel B could be performed by the amplifier so that both diodes would be in the same direction.

Whatever the circuit configuration, the basic operating criteria remain the same as demonstrated by the following two specific cases.

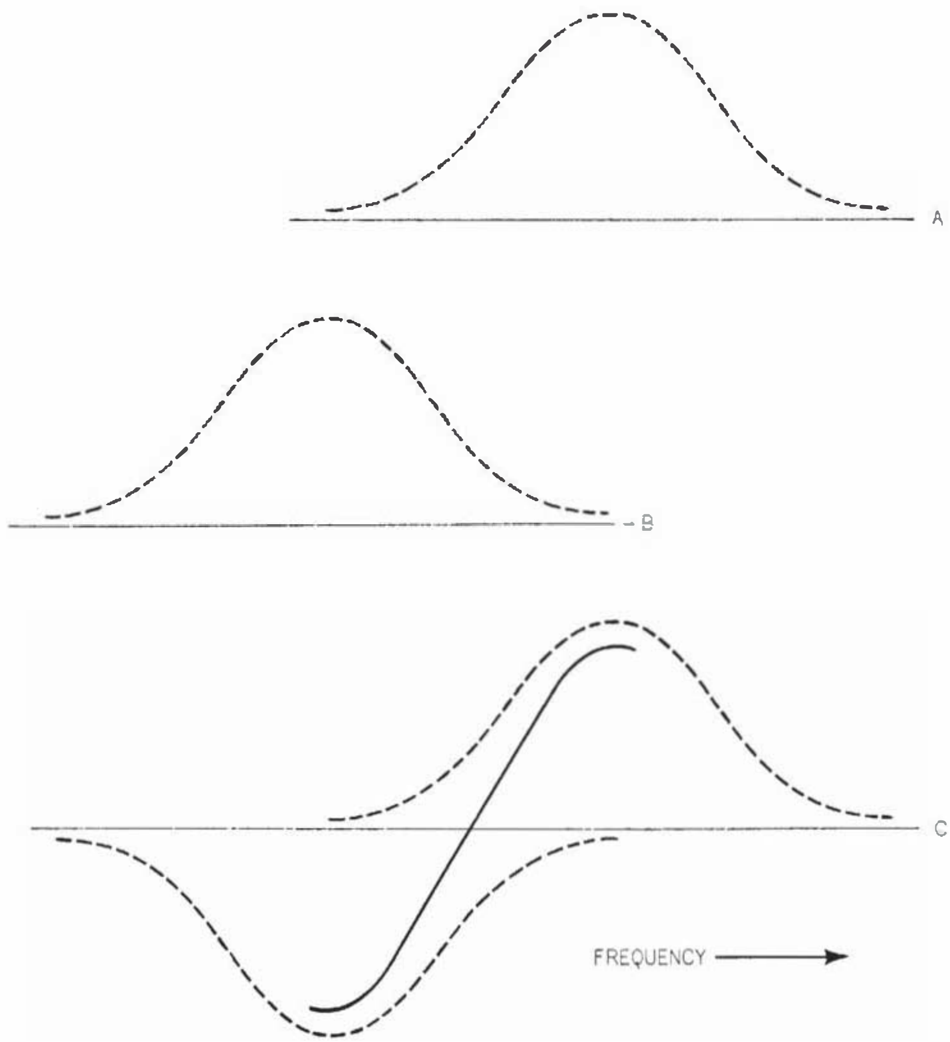


Fig. 2-4. LC-discriminator frequency characteristics.

TRANSMISSION-LINE DISCRIMINATOR

Transmission lines behave as resonant circuits as discussed previously. This means that transmission lines can be used as the resonant elements in a discriminator. The basic advantages of using transmission lines are the standard ones of wide bandwidth and/or high frequency. Figs. 2-5 and 2-6 are a schematic and frequency characteristic of a transmission-line discriminator.

equivalent reactance

The voltage-frequency relationships shown in Fig. 2-6 are best understood with reference to the transmission-line impedance diagrams illustrated in Figs. 1-4 and 1-5. It will be observed that a short-circuited line less than $\frac{\lambda}{4}$ in length represents an inductive reactance whose value increases with increasing line length, while an open transmission line less than $\frac{\lambda}{4}$ long is equivalent to a capacitive reactance decreasing with increasing line length. Electrical line length and frequency are, of course, related by $f = \frac{c}{\lambda}$, where: f is frequency, λ is wavelength, and c is the propagation velocity of electromagnetic radiation which is 3×10^8 meters/second in vacuum. In other words, the wavelength, λ, decreases with increasing frequency, which means that the electrical length of a line (in wavelengths) increases with increasing frequency for a fixed length of line (in inches, meters, etc.). The voltage-frequency relationships of Fig. 2-6 are now evident; the shorted line represents an inductance, the voltage appearing across this line should therefore increase with frequency, while the open line represents a capacitance and the voltage across it should decrease with frequency. After polarity inversion and summation we end up with Fig. 2-6C.

electrical-line length

broad-band discriminator

As previously indicated, one of the advantages of transmission-line discriminators is their inherently broad bandwidth. This can best be understood with reference to Fig. 2-6C. The full operating range shown in Fig. 2-6C is $\frac{\lambda}{4}$, with the zero voltage crossover in the center of the range at $\frac{\lambda}{8}$. The discriminator cannot be operated over this entire range as the response becomes nonlinear near the edges. However, even with a restricted range, the

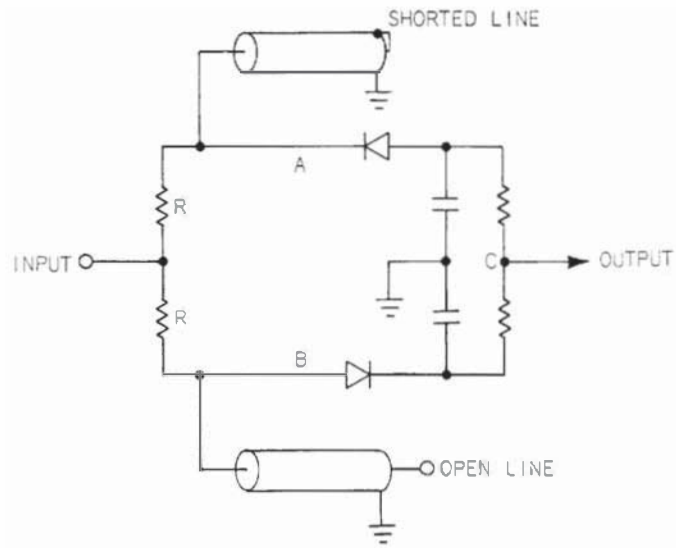


Fig. 2-5. Transmission-line discriminator.

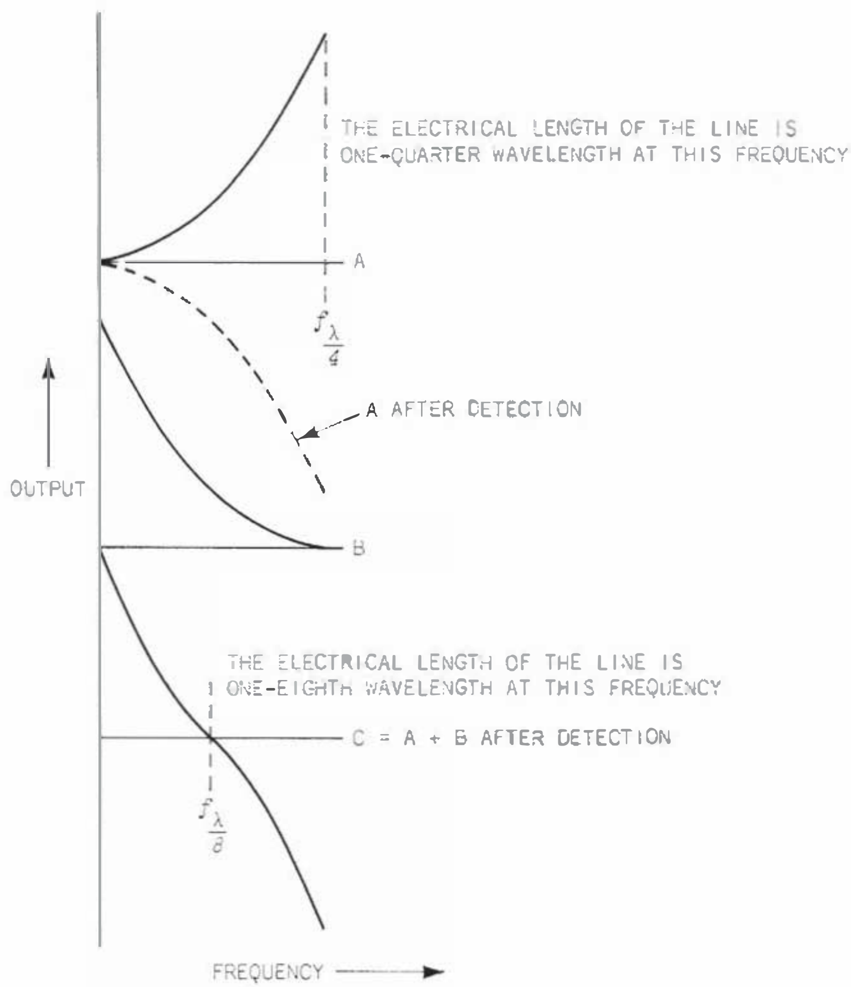


Fig. 2-6. Transmission-line discriminator frequency characteristics.

bandwidth of this type of discriminator comes out extremely wide. For example, assume that we will operate the discriminator within $\frac{\lambda}{16}$ from the center, or from $\frac{\lambda}{16}$ to $\frac{3\lambda}{16}$. This means that the total operating range is $\frac{\lambda}{8}$, which is equal to the center point of the range; the bandwidth of the discriminator is equal to the center frequency. Transmission-line discriminators having a 100-MHz bandwidth at a 100-MHz center frequency, or a 2-MHz bandwidth at a 2-MHz center frequency, are therefore, practical. Note that best linearity is obtained when the input resistor, R, is equal to the transmission-line characteristic impedance, Z_0.

operating range

It was noted that while relatively low-frequency (in the MHz region) broad-band transmission-line discriminators are practical, they present a special problem. Note that the free-space wavelength at 1 MHz is almost 1,000 ft. Even assuming considerable foreshortening of the line due to the use of dielectric, an eighth of a wavelength is still a tremendously long line. As a result transmission-line discriminators at these frequencies frequently utilize artificial foreshortening of the lines by the addition of shunt capacitors across the lines. The effect of the shunt capacitors is shown in Fig. 2-7.

fore-shortened lines

It will be observed that for the unforeshortened lines, reactances (one inductive and the other capacitive) are equal at $\frac{\lambda}{8}$. Adding shunt capacity to a capacitive reactance will result in a reduced

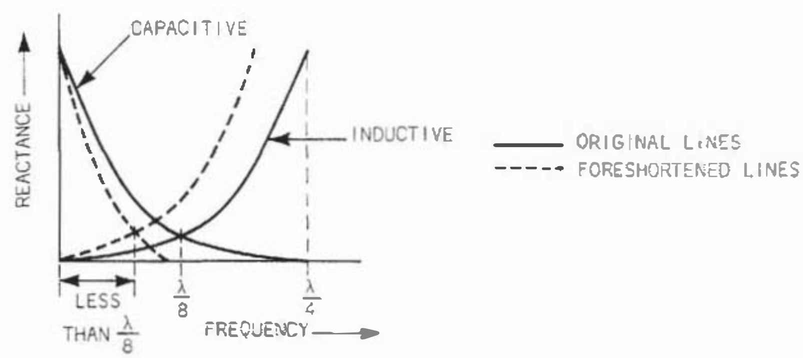


Fig. 2-7. Reactance vs frequency of foreshortened transmission lines.

reactance (capacitors in parallel) at all frequencies except at zero where the reactance is infinite. Adding shunt capacity to an inductive reactance increases the total reactance (think of a tank circuit closer to resonance) at frequencies below resonance except at zero where the inductance represents a short circuit. The dashed lines in Fig. 2-7 show the foreshortened reactances, which will be observed to be equal at a frequency where the original lines are less than $\frac{\lambda}{8}$ long. Hence the length of line used is reduced by the addition of these shunt capacitors. Another foreshortening technique is to add inductance to the shorted line and capacitance to the open line. Such foreshortening techniques can reduce the length of the lines by up to 20%. The addition of foreshortening elements, however, reduces the range of linear operation, since the curvature in the output-voltage-versus-frequency curve is increased. One important effect of using shorter transmission lines is a reduction in harmonic nonlinearities.

CRYSTAL DISCRIMINATOR

quartz crystal

As discussed later in more detail, the quartz crystal has unique properties as a resonator. The extreme frequency stability and high Q make the quartz crystal ideally suited to the design of high-stability narrowband discriminators. As discussed in the section on quartz crystals, the prime resonance is usually in the series mode, necessitating a slight variation from the basic discriminator circuit previously discussed. A complete crystal discriminator in schematic form is shown in Fig. 2-8.

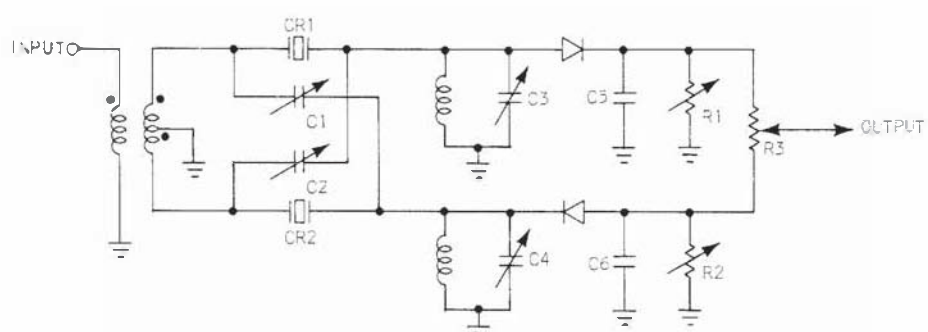


Fig. 2-8. Detailed schematic of crystal discriminator.

neutralize crystal capacitance

The input transformer splits the incoming signal into two channels with minimal power loss. Capacitors C1 and C2 provide neutralizing paths, thus reducing the effect of the crystal's shunt capacity. Neutralizing techniques are discussed in more detail in the section on crystals. Crystals CR1 and CR2 are series resonant at the band edges of the discriminator respectively, with CR1 usually at the higher frequency. These crystals operate into the high-impedance tank circuits which are tuned to the crystal frequencies by means of C3 and C4 respectively. Thus, even though LC tank circuits are used as in a conventional discriminator, the peak frequencies are firmly established by the crystals. The rest of the discriminator shows a conventional design including all the adjustments.

The setting of R3 establishes the balance point between the two channels, resistors R1 and R2 affect both the balance and the sensitivity (volts out per hertz in), and in conjunction with capacitors C5 and C6, R1, R2, and R3 form a low-pass output filter.

crystal high Q

Because of the high Q and therefore sharp skirts of the crystal resonators, crystal discriminators usually have high sensitivity and narrow bandwidth (less than 1% of center frequency).

TRANSMISSION-LINE TRANSFORMERS

A) Theory of Operation

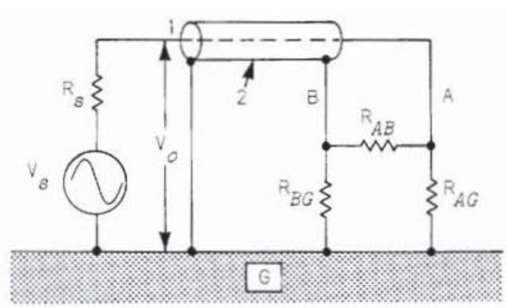
There are very few modern electronic instruments that don't contain at least one transformer. This transformer may be in the power supply, or an output driver, or perform interstage coupling in an IF or filter system, or perform a myriad of other functions. Typical transformer functions include voltage step-up or step-down, low-power-loss impedance matching, signal-polarity inversion and signal splitting. Fig. 2-8 in the discriminator section shows a transformer being used to split a single-ended input into two balanced channels. A conventional transformer is perfectly suitable to this application, since a crystal discriminator is a relatively narrowband device.

advantage

Standard transformers are in general not suitable for very wideband and/or relatively high-frequency (hundreds of MHz) applications. This is because the interwinding capacitance and leakage inductance introduce resonance peaks. It is, therefore, necessary to use a transformer where the parameters are distributed and where the stray capacitance and inductance is absorbed as part of the general structure. The transmission-line transformer is such a device.

coaxial-transformer analysis

Consider a coaxial transmission line suspended above a ground plane as shown in Fig. 2-9. Assuming that the line is terminated in the characteristic impedance (i.e., $R_S = R_{AB} = Z_{1,2}$ the characteristic impedance of the line consisting of conductors 1 and 2, with R_{BG} and R_{AG} large), then from transmission-line theory we have: $V_O = \frac{V_S}{2}$, and A would be at $+0.5\ V_O$ and B at $-0.5\ V_O$, since in the absence of a ground current the current in R_{AG} must equal that of R_{BG} and be of opposite polarity. Thus, in the absence of ground current, a single ended input is converted into a double-ended output. Unfortunately R_{BG} is shunted by the relatively small impedance $Z_{2,G}$ (remember we have assumed R_{AG} and R_{BG} large) so that V_B is less than V_A, the output is not balanced and a ground current flows.



SUBSCRIPTS		DESCRIPTION
1	--	CENTER CONDUCTOR
2	--	SHIELD
A	--	OUTPUT
B	--	OUTPUT
G	--	GROUND
$Z_{1,2}$	--	CHARACTERISTIC IMPEDANCE OF LINE.
$Z_{1,G}$	--	CHARACTERISTIC IMPEDANCE OF LINE COMPOSED OF CENTER CONDUCTOR AND GROUND PLANE.
$Z_{2,G}$	--	CHARACTERISTIC IMPEDANCE OF LINE COMPOSED OF SHIELD AND GROUND PLANE.

Fig. 2-9. Transmission line above a ground plane.

balance

From the foregoing it will be observed that we have two options for making a better balance, one is to increase $Z_{2,G}$ and the other is to balance both sides of the line with respect to ground. The value of $Z_{2,G}$ cannot be increased without limit since $Z_{2,G}$ is a function of the physical separation between the transmission line and ground plane, and the line length.

The second option must therefore be resorted to. A simple system of this type is a two-wire transmission line above a ground plane, with the wires balanced with respect to ground. Here the output will be balanced, the degree of balance being determined by the values of R_{AG}, R_{BG} and $Z_{2,G}$.

Thus, a single-ended input is converted to a balanced output by properly adjusting the impedance from each side of the line to ground. It is obvious that since $Z_{1,G}$ and $Z_{2,G}$ are not infinite, current must flow in the ground plane.

primary mode

common mode

In the above system, two waves are caused to travel down the line. One wave travels between lines 1 and 2 and its behavior is determined by the characteristics of $Z_{1,2}$, and the other is sustained by the ground current as determined by $Z_{1,G}$ and $Z_{2,G}$. The wave propagating down $Z_{1,2}$ contains the desired information and is therefore the primary mode. The ground wave is generally undesirable and is referred to variously as the "common mode" or "secondary mode." It should be noted that although the secondary mode causes all kinds of difficulties, transmission-line transformers work by virtue of the existence of the secondary mode. We could not, for example, obtain a balanced output except through the excitation of a secondary mode.

B) Practical Configurations

impedance transformer

So far we have considered one of the transformer functions: The single ended-to-balanced configuration. Transformers are, of course, used for other functions, the most frequent of these being the impedance transformer. Impedance transformation in transmission lines is obtained by the proper connection of a number of lines in various series-parallel configurations. For example, a length of 50-Ω line connected to a 50-Ω source and 50-Ω load is essentially a one-to-one impedance transformer. If 50-Ω transmission line is not available, then two equal lengths of 100-Ω line connected in parallel will serve the purpose. Unequal impedances are matched by combining transmission lines in series and parallel as shown in Fig. 2-10.

Here two equal-length 50-Ω lines are connected in series at one end (100 Ω) and in parallel on the other end (25 Ω). You will note that by series connection it is not meant a cascade of two transmission lines where center conductors are connected together and shields together, but rather a reversal of center conductor to shield.

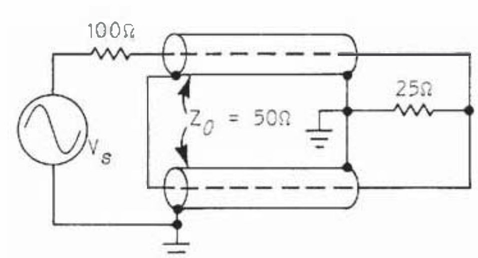


Fig. 2-10. Four-to-one transmission-line impedance transformer.

From the foregoing it is clear that a judicious choice of line impedance, line length and line series-parallel connection results in an ability to match almost any two impedances. This is fine theoretically, but as a practical matter there are difficulties. These arise mainly from the conflicting requirements that the transmission lines should be very short so as to obtain minimum delay and distortion. The secondary-mode (common-mode) amplitude across the output impedances should be small or/and be delayed a long time with respect to the primary mode. It is also important that both sides of the transmission line be well balanced with respect to ground all along the length of the line. Fortunately, there is a simple configuration that fulfills all of these requirements. This structure is the multifilar winding on a ferrite toroid. An example of a bifilar winding (2 wires forming a single transmission line) and the equivalent circuit for a polarity-reversing transformer is shown in Fig. 2-11.

delay and distortion

balance

ferrite-core toroid

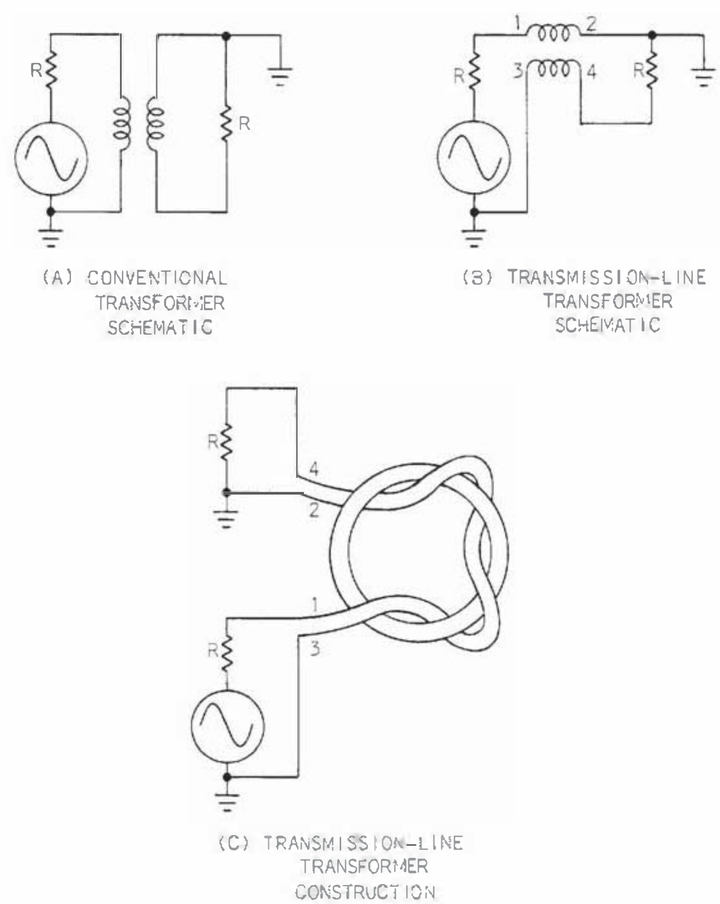


Fig. 2-11. Polarity-reversing transformer.

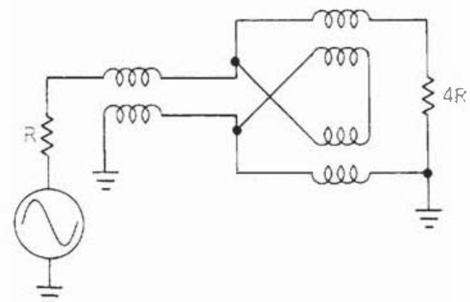


Fig. 2-12. Transmission-line transformer, four-to-one unbalanced configuration.

The toroidal construction meets all of the major requirements of transmission-line transformers. This construction lends itself to small physical size, and the interposition of ferrite between the transmission line and chassis ground minimizes effects due to the secondary mode. Further, since both transmission-line wires are wound as a pair (bifilar), both sides of the line will be balanced to ground regardless of the physical layout or mounting of the transformer.

Fig. 2-11 shows a simple configuration; more complicated configurations are, of course, possible. Fig. 2-12 shows a schematic representation of a relatively complex 4:1 unbalanced configuration.

input isolation

This is quite similar to the configuration shown in Fig. 2-10 except for the addition of a third transformer at the input. This transformer provides input isolation and a balanced drive to the series (input) parallel (output) connection of the 4:1 impedance transformer. There are several ways of constructing this configuration. For example, all three transformers on one core, three separate cores, etc.; further the input transformer is often absorbed as part of the others so that only two transformers are actually used. There are several obvious differences and some not so obvious differences between transmission-line transformers and ordinary transformers. An obvious difference is that transmission-line transformers do not provide DC isolation, whereas ordinary transformers do. This means, for example, that unusual bias techniques must be utilized when using transmission-line transformers in active circuits. Not so obvious (at least from looking at a schematic

no DC isolation

representation) is the fact that these transformers are really transmission lines that perform transformer functions and not vice-versa. It is quite a tricky matter to arrive at the right conclusions regarding the performance of a transmission-line transformer when treating it as if it were an ordinary transformer connected in some peculiar manner. The transmission-line nature of these devices must be kept in mind at all times for a proper understanding of how they work.

CRYSTAL-DIODE CHARACTERISTICS

Crystal diodes, also known as crystals or diodes, are used extensively in spectrum-analyzer circuits. Such circuits include mixers, modulators, limiters, pulse generators, detectors, attenuators, signal shapers, and others. Some of these circuits operate by virtue of some unique diode parameter available in specialized diodes only. An example of such an application would be pulse generation by means of the step-recovery diode. These unique and specialized diode parameters will be discussed as needed in conjunction with specific circuit applications. In this section we shall concentrate on a discussion of the general nature of microwave diodes.

Fig. 2-13 is an equivalent circuit of a semiconductor diode. The series resistance R_s and barrier capacitance C_b are undesirable parasitic elements which are minimized as much as possible. The barrier resistance R_b is nonlinear, the value being dependent upon the applied voltage. For most practical applications this is usually described by---

$$i = I\,(e^{\alpha V} - 1) \qquad \text{(Eq 1)}$$

where: α = a constant defining curvature of the diode characteristic

V = volts across diode

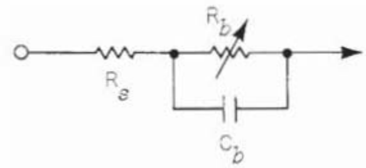


Fig. 2-13. Semiconductor-diode equivalent circuit.

Equation 1 tells us a great deal about the behavior of the diode under various circumstances. For example, using a series expansion we observe that Equation 1 is equivalent to:

$$i = \frac{I\alpha}{1!} V + \frac{I\alpha^2}{2!} V^2 \ldots + \frac{I\alpha^n}{n!} V^n \qquad \text{(Eq 2)}$$

From Equation 2 it is apparent that because of the higher-order terms of V, the diode can be used as a frequency multiplier (mixer, modulator applications).

diode as a nonlinear resistance

These equations lead to a theory of how diodes can best be used as nonlinear resistance elements in various applications. For example, if we apply two signals to the diode,

$$V = A \sin a + B \sin b \qquad \text{(Eq 3)}$$

Substituting Equation 3 into Equation 2 we have:

$$i = A[K_1 + 3K_3(\frac{1}{4} A^2 + \frac{1}{2} B^2) + 5K_5(\frac{3}{4} A^2B^2 + \frac{3}{8} B^4$$

$$+ \frac{1}{8} A^4)] \sin a - A^2[\frac{K_2}{2} + K_4(\frac{1}{2} A^2 + \frac{1}{2} B^2)] \cos 2a$$

$$- A^3[\frac{1}{4} K_3 + 5K_5(\frac{1}{4} B^2 + \frac{1}{16} A^2)] \sin 3a, \text{ etc.} \qquad \text{(Eq 4)}$$

$$\text{where: } K_n = \frac{I\alpha^n}{n!}$$

It will be observed that to get a large second-harmonic output (cos 2a) it is desirable to maximize K_2 and K_4. Similar computations can be made for mixing where, for example, the coefficient of cos (b - a) is $AB[K_2 + 3K_4(\frac{1}{4} A^2 + \frac{1}{2} B^2)]$ and again K_2 and K_4 would be optimized for maximum output at the frequency (b - a).

diode switch

In spite of these possibilities it should be recognized that Equation 1 does not tell the whole story. A diode can be used not only as a nonlinear resistance which can be expressed by an infinite series, but also as a switch which cannot be expressed as a simple series. Many applications, including low-loss mixers and modulators, are best understood when the diode is considered as a switch. An ideal switch presents zero resistance in the closed position and infinite resistance in the open position. Since neither zero resistance nor infinite resistance dissipate power at finite currents and voltages, it is obvious that such a configuration will exhibit extremely low losses provided other sources of loss can be minimized.

ideal switch

PHASE DETECTORS AND SAMPLERS

phase detector

A phase detector is a device which produces an output which is proportional to the phase difference between two signals. Phase detectors are commonly used in spectrum-analyzer oscillator phase-lock systems. In general, because of the high frequencies and wide bandwidths involved, spectrum-analyzer phase detectors are usually of the sampling type.

Theoretically there are many circuit configurations that will behave as phase detectors. For example, an ideal signal multiplier and appropriate filter will produce an output which is proportional to the phase difference between the two input signals. In spectrum analyzers the basic phase detector is usually of the balanced diode-peak detector variety. Fig. 2-14 is a simplified balanced-peak-detector circuit. In practice the reference signal e_r and the signal to be operated on, e_s, are injected by transformers or other means that permit good balance.

balanced-diode peak detector

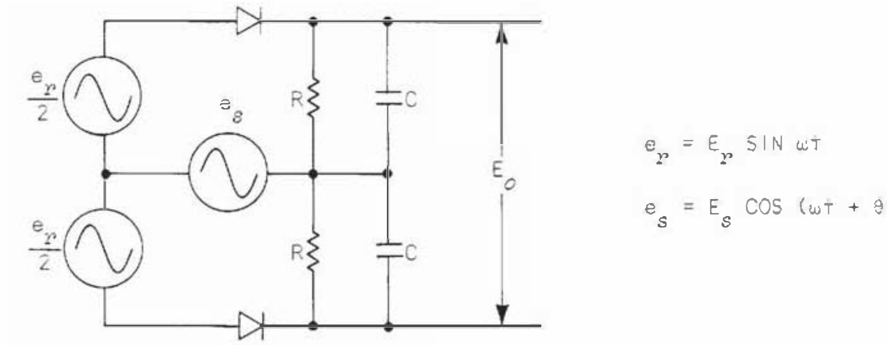


Fig. 2-14. Balanced-diode phase detector.

detection range

It can be shown that if E_r is large compared to E_s then $E_o \approx 2E_s \sin\theta$. This is a practical circuit, though it has some drawbacks. First, the output E_o is not proportional to the phase difference, neglecting the fixed 90° difference between (sin ωt) and (cos ωt), but rather to sin θ. Second, the output E_o is proportional to the signal level E_s. This means that E_s must be constant if E_o is to be a function of θ only. Fortunately in spectrum-analyzer circuits the steady state condition for the phase detector is that which leads to $E_o = 0$, which occurs at $\theta = 0°$. Thus, e_s can be permitted to fluctuate between the level which is sufficient to overcome system noise (good signal to noise ratio) and the level which would cause diode breakdown. Maximum phase-angle detection range is ±90° where E_o goes to a maximum of $2E_s$.

Another way of considering the circuit in Fig. 2-14 is this: The sum of signal and reference voltage is applied to one diode while their difference is applied to the other diode. The detected output is taken across both diode load resistors. If the load-resistor R is substantially greater than the diode series resistance, and the output time constant (RC) is long compared to one cycle, we have essentially a peak detector with a total output of $E_s - \frac{E_r}{2} + E_s + \frac{E_r}{2} = 2E_s$. What we have described is an arrangement of two properly phased peak detectors.

Now let us consider in more detail how a peak detector operates.

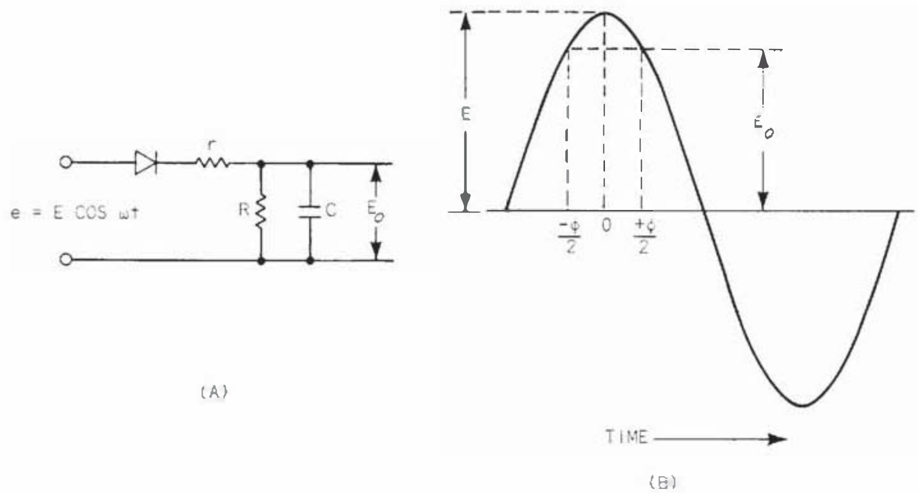


Fig. 2-15. Peak detector and associated waveform.

Fig. 2-15 shows an ordinary peak detector and the associated waveform, where r is the forward resistance of the diode, and RC, the output time constant, is large compared to the period $1/f$.

From Fig. 2-15B we observe that E_o is a function of E and $\frac{\phi}{2}$, namely $E_o = E \cos \frac{\phi}{2}$, where $\frac{\phi}{2}$ is dependent on the ratio $\frac{r}{R}$. In other words for a theoretically perfect peak detector $\frac{r}{R} \to 0$, $\frac{\phi}{2} \to 0$, so that $E_o \to E$. A good peak detector has a DC output (E_o) nearly equal to the peak AC input (E), and the diode conducts for a very short time, just enough to replenish the charge lost due to the noninfinite nature of the product RC.

peak-detector analysis

So far we have considered a standard two-diode balanced phase detector where both the signal (e_s) and the reference (e_r) are considered to be sinusoidal. Further, it was shown that the diodes in the detector have a very small conduction angle. It, therefore, follows that the reference signal need not be sinusoidal, provided it is of a form causing the detector diodes to conduct for the necessary angle. Such a signal is a narrow pulse that turns the detector diodes on for a short time interval. The system just described (where e_r is in the form of a narrow pulse) constitutes a sampling type of phase detector.

narrow-pulse reference

sampling phase detector

sampling phase-detector advantage

Sampling phase-detector systems are generally more complex than nonsampling systems requiring, for example, narrow-pulse generators and associated wideband-coupling circuitry. The sampling type of detector, however, has characteristics which make it ideal for wideband spectrum-analyzer systems. The nonsampling phase detector requires a stable reference at the same frequency as the signal. This means that in the case of wideband spectrum analyzers, where the local oscillator is tuneable over a wide frequency range, the reference must be tuneable over a wide frequency range. But stable tuneable reference signals are hard to come by. Further, the sampling phase detector will work with the reference set to a subharmonic of the signal frequency, whereas in a nonsampling system this is difficult to do. For these and other reasons sampling phase detectors predominate in spectrum-analyzer phaselock circuits.

frequency-domain analysis

The sampling phase detector, with reference set to a subharmonic of the signal, can be considered from two points of view. These are frequency-domain analysis, and time-domain analysis. Frequency-Domain Analysis: Consider a signal (e_s) at a frequency f and a reference pulse train at a frequency f/n with each pulse V-volts high and T-seconds wide.

From Fourier theory we can consider the pulse train as consisting of an infinite number of sinewaves, where the amplitude of the n^{th}-harmonic C_n at frequency f the same as e_s is given by:

$$C_n = \frac{2V\tau f}{n}\left(\frac{\sin \pi\tau f}{\pi\tau f}\right).$$

For the case where $\tau f << 1$ this becomes $C_n \approx \frac{2V\tau f}{n}$.

Conceptually we are now back to the original circuit where the signal e_s is compared to a reference e_r at the same frequency as e_s. This is fine for getting an intuitive feel for how the circuit operates but it does not tell the whole story. In particular there is the question of what do the other harmonics do to the circuit, and we originally assumed a large E_r but the new reference signal is small since it goes as 1/n.

time-domain analysis

Time-Domain Analysis: Consider a diode peak detector as shown in Fig. 2-15A. The diode is biased to present a high impedance to the signal; most of the signal appears across the diode and E_o goes to zero. If we now pulse the diode on for a short interval during each cycle of e_s the result is a waveform similar to Fig. 2-15B. Fig. 2-15B shows that E_o will vary from zero for the case where the turn-on pulses are synchronized with the zero crossing of e_s, to some maximum level when the turn-on pulses are synchronized with the peak of e_s. This is precisely what a phase detector is supposed to do: Produce an output which is a function of the phase difference between two signals.

output varies with phase difference

A one-to-one correspondence between the reference-pulse frequency (turn-on pulses) and the signal e_s is fine if the signal frequency is fixed. If, however, the signal frequency is variable we have the problem of producing a variable-reference source. This is generally not practical. A simple solution is to operate the reference at a frequency which is a subharmonic of the signal frequency. This, of course, limits the rate at which change of relative phase information can be acquired. For example, using a 1-MHz reference source the change of relative phase information is updated every microsecond regardless of the period of the signal e_s. Fortunately this does not present any problem in spectrum analyzers. A second problem is that the signal (e_s) can be compared to the reference (e_r) at discrete frequencies only. Again if we use a 1-MHz reference, then the signal can be compared to the reference in 1-MHz steps only. This can be taken care of in two ways. One is to use a frequency offset in another oscillator in the spectrum-analyzer superheterodyne chain. We do this quite often; the offset control is called IF Center-Frequency control. A second solution that works well, especially when the reference frequency is quite low in comparison to the signal frequency, is to make the reference frequency slightly adjustable (e.g., 0.1%). For example, if the signal goes 1 to 2 GHz and the reference is 1 MHz then tuning the reference by 1 kHz permits continuous coverage. This follows since we are in effect comparing 1 GHz to the 1000^{th} harmonic of 1 MHz and 1000 times 1.001 MHz = 1 GHz + 1 MHz where the thousand-and-first harmonic of 1 MHz can take over.

change of relative phase information

adjustable reference frequency

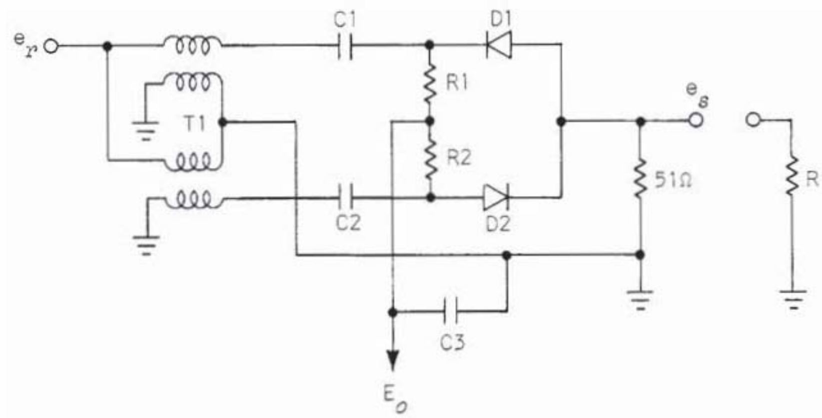


Fig. 2-16. Sampling phase detector.

typical phase detector

Fig. 2-16 shows a typical sampling type of phase detector. The single-ended source of reference pulses is converted into a balanced drive by means of the wideband transmission-line transformer (T1); capacitors C1 and C2 provide DC isolation, the 51-Ω resistor provides a proper termination for the source of e_s, diodes D1 and D2 are off (this can be accomplished by self bias as in Fig. 2-16 or external bias) until turned on by e_r pulse, and R1 and R2 in conjunction with C3 form the main memory circuit for E_o (C1, C2 and the load R affect the memory characteristics as well).

NARROW-PULSE GENERATORS

applications

Narrow-pulse generators, or impulse generators as they are sometimes called, find frequent use in spectrum-analyzer circuits. Applications include picket-fence generation for marker or oscillator use and strobe generators for sampling phase detectors. It is not the purpose here to discuss more or less ordinary pulse-generator systems such as multivibrators or blocking oscillators, though these are on occasion used in spectrum analyzers. We shall restrict our discussion to pulse-forming transmission lines, step-recovery diodes, and avalanche transistors.

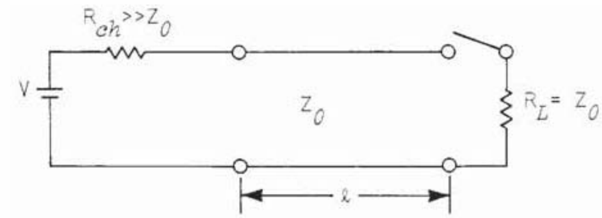


Fig. 2-17. Pulse-forming transmission line.

A) Pulse-Forming Transmission-Line Circuits.

pulse-forming transmission lines

One of the best known and, at least in theory, simplest narrow-pulse-generating circuits is the open-circuited transmission line as shown in Fig. 2-17.

charging the line

The open transmission line is connected to a DC voltage through a charging resistor (R_{ch}). In time the transmission line (which behaves essentially like a small capacitor) charges up to V, the charging voltage. The line is now terminated by a load resistor (R_L) which is equal to the characteristic impedance (Z_0). What happens -- at the instant of switch closure, we have in effect connected a resistor R_L to a voltage-source V having an internal-impedance $Z_0 = R_L$. As a result, by ordinary resistive voltage division, the voltage across R_L is V/2. But, the voltage at the other end of the line is still V, since the disturbance caused by the connection of R_L to the line cannot instantaneously propagate down the line. The disturbance caused by the connection of R_L to the line is such as to cause the voltage on the line to change from V to V/2. In other words, connecting R_L across the line causes a wave of amplitude -V/2 to propagate down the line from the load to the source. When this propagating wave reaches the charging resistor (R_{ch}) it gets reflected essentially from an open circuit, since R_{ch} is very large, with a reflection coefficient of

$$\Gamma = \frac{Z_{ch} - Z_0}{Z_{ch} + Z_0} \approx 1.$$

generating the pulse

The voltage wave of -V/2 now travels back toward the load leaving zero volts behind it since V/2 + (-V/2) = 0. When this reflected wave reaches the load the voltage across R goes to zero. Thus, we

have generated a pulse of amplitude V/2 and pulse-width $T = 2\ell/v$ where ℓ is the length of the transmission line and v is the propagation velocity down the line. For example, for air line (v = c, the speed of light in vacuum) we get T = 0.5 nanoseconds, for ℓ = 7.5 cm.

switch limitations

This type of pulse generator has been fairly popular since it lends itself to the generation of large yet narrow pulses. Several hundred volts at less than a nanosecond wide can be achieved. The major disadvantage is the limitations of the switch. Since the load, R_L, must be connected directly to the line, it is necessary that the switch be either small or form a constant Z_0 transmission line. Without going into all the problems, it should be noted that switching difficulties preclude the use of this type of pulse generator at high repetition rates much in excess of 1 kHz and it is seldom used in spectrum-analyzer circuits.

B) The Avalanche Transistor

avalanche characteristics

The avalanche transistor is a useful device for generating fast turn-on, high-power pulses. For example, 50 volts at one ampere with nanosecond pulse widths can be generated. This is accomplished by switching the transistor between two modes of operation as illustrated by the avalanche-breakdown characteristic, Fig. 2-18. In its normal mode of operation the transistor operates in region 1. At a collector current of I_A, which is reached by biasing the base-emitter junction forward (i.e., control of base current), the transistor switches into the avalanche mode of operation represented by region 2. When the current

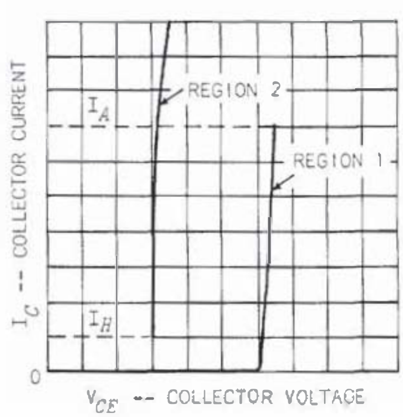


Fig. 2-18. Avalanche breakdown characteristic.

avalanche switching

has dropped to I_H, the transistor turns off and switches back to region 1. Refer to Fig. 2-19. Normally the transistor is off. The positive trigger pulse on the base turns it on. The transistor switches into the avalanche mode, and the dynamic impedance drops to a low value (e.g., 30 Ω). The capacitor discharges through R_L and the transistor, until the current drops below I_H when the transistor switches back into region 1. The transistor is made to stay in region 1 until the next turn-on pulse by making the limiting-resistor R_{lim} large enough to limit the current below I_H, i.e., large enough to allow switching out of region 2 and return to region 1. Fig. 2-19B shows typical waveforms.

C) The Step-Recovery (Snap-Off) Diode

In an ordinary diode, the reverse current usually recovers slowly to zero when the diode is switched from forward to reverse bias. In the step-recovery diode (SRD), the diode characteristics are optimized to reduce the reverse-current recovery time while increasing storage time, resulting in a fast recovery step. The faster the recovery time with respect to diode-storage time the better the diode. The fast change in current can be used to generate narrow pulses by such means as pulse clipping, driving an inductance ($V = L\,\frac{di}{dt}$), or other means.

The fast-pulse generation capability of this diode is due to the fact that the SRD looks like a large capacitor in series with a resistor when biased in the forward direction. As a result the diode stores charge when biased ON. At reverse bias the charge is depleted during the storage time, generating a narrow pulse when all the charge is gone. The capacitive nature of this diode, as shown in Fig. 2-20, has led various people to consider the SRD as a form of varactor.

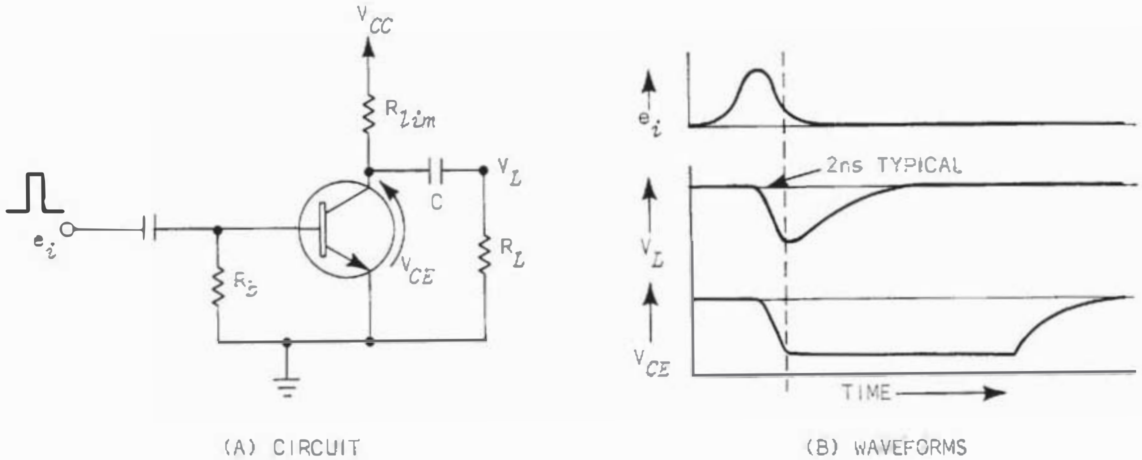


Fig. 2-19. Avalanche transistor.

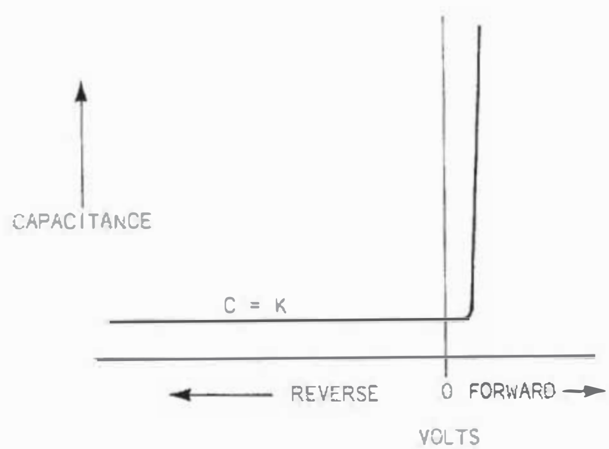


Fig. 2-20. Ideal SRD capacitance characteristic.

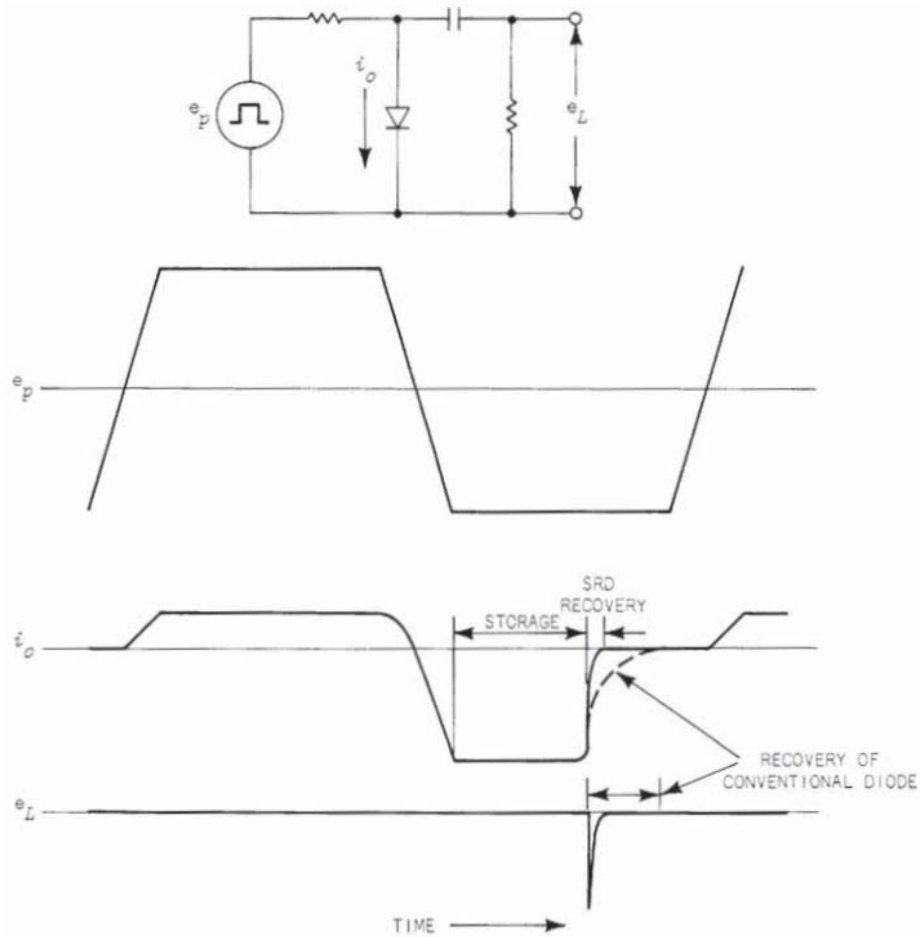


Fig. 2-21. Step-recovery diode circuit and typical waveforms.

Step-recovery pulse generators require a driving source, with the SRD pulses occurring in synchronism with the driving-source rate. The advantage of the SRD system is that a wide-pulse large-risetime source is converted to a narrow-pulse (e.g., 0.1 ns, 10 V into 50 Ω) fast-risetime source. A simplified diagram of a step-recovery pulse generator is shown in Fig. 2-21 with an illustration of a cycle of operation.

YIG RESONATORS

Single-crystal Yttrium Iron Garnet (YIG) forms a high-Q microwave resonator when fabricated into highly polished elements such as spheres or discs in the order of .01 inches in diameter. YIG elements are of great interest because these possess the desirable property of electronic tuning. Basically the behavior of YIG is best explained by the fundamental properties of a ferromagnetic material. In a ferromagnetic material the molecules of the crystal possess a net magnetic moment. These magnetic moments which are normally distributed in a random manner, can be aligned in one direction by the application of an external static magnetic field. If an alternating magnetic field of appropriate frequency is now applied at right angles to the static field, the crystal magnetic moment will precess around the static field, where the rate of precession is determined by the properties of the YIG material and the magnitude of the static field. In other words, *the YIG crystal exhibits the property of resonance,* where the resonant frequency is determined by the characteristics of the material and the magnitude of the applied external magnetic field.

realign magnetic moments

moment precessing

Since the resonant frequency is determined by the magnitude of the applied static magnetic field, it follows that a YIG resonator can be tuned electronically by controlling the current in the electromagnet. Further, since the resonant frequency is directly proportional to the external field strength, a YIG resonator has a theoretically linear tuning characteristic as a function of applied current.

resonance tuned by electro-magnetic field

Linear electronic tuning together with wide frequency ranges and relatively high-frequency (well above 10 GHz) operation make YIG resonators desirable components. There are, however, many problems in the application of YIG in practical circuits.

Many of the problems in using YIG stem from temperature effects. This is in addition to the basic problem of obtaining good spheres (i.e., uniform material of proper purity fashioned into highly spherical well-polished balls).

equation for resonance

small precision spheres

The following equation, which may be ignored without loss of continuity, provides the physical basis for the subsequent discussion of the YIG sphere:

$$f_r = \gamma\ [H_0 \pm H_a + (N_T - N_Z)\ 4\pi M_s]$$

where: f_r = resonant frequency

H_0 = applied tuning field

H_a = Anisotropy field in crystal. This field is generated by the crystal, and adds to H_0. This field (about 48 oersted in YIG at room temperature) is a function both of temperature and direction (hence name anisotropy - direction dependent).

N_Z & N_T - Demagnetizing factors in the direction of H_0, and transverse to H_0 respectively. For a sphere $N_T = N_Z = 1/3$.

$4\pi M_s$ -- The saturation magnetization of the material. $4\pi M_s$ is affected by temperature. For proper operation the external field, H_0, must be greater than $4\pi M_s$, which depending on the material can be about 400 to 1800 gauss.

γ -- Gyromagnetic ratio for electron (2.8 MHz/oersted).

A proper YIG resonator design must take all of the above into account. Using highly spherical balls will minimize $4\pi M_s$ effects since $N_T = N_Z$ for a perfect sphere. Further saturation magnetization effects are reduced by using the lowest $4\pi M_s$ material and smallest spheres practicable. This is, however, limited by increased difficulty in coupling to the YIG ball. The effects of the anisotropy field are reduced by aligning the sphere with respect to the direction of H_0 so as to minimize H_a temperature effects. There are many other problems associated with the design of YIG circuits. One cannot, for example, exchange spheres between two working circuits, in the manner of exchanging transistors, and expect everything to work without any further alignment.

YIG resonators can be used in all the circuits where resonators can be used. Such circuits include filters, oscillators, discriminators, etc.

YIG coupling circuit

Fig. 2-22 shows a typical YIG coupling circuit. It should be noted that the load, and tuning-field H_0, are at right angles to each other. If a second external circuit were connected to the YIG sphere, such as in a filter, the second loop would be set at right angles not only to H_0 but to the first loop as well. This reduces coupling between the loops except for that provided by the resonant YIG sphere.

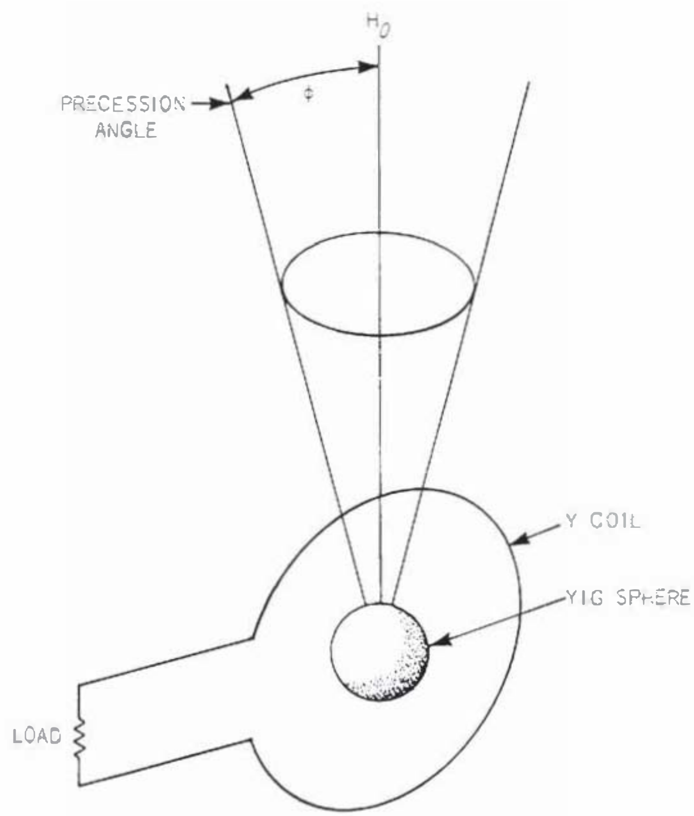


Fig. 2-22. Basic relationships for YIG-resonator design.

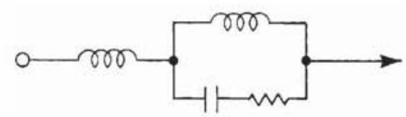


Fig. 2-23. YIG-resonator equivalent circuit.

Fig. 2-23 shows the equivalent circuit of a YIG sphere. Note the similarity between Figs. 2-23 and 2-24, the equivalent circuit of a quartz crystal.

QUARTZ CRYSTALS

Quartz-crystal resonators exhibit extremely high Q's, typically from 10,000 to above 100,000. In addition quartz-crystal resonators exhibit good temperature stability over wide temperature ranges, low change as a function of time, and other desirable features such as relatively low microphonics and good behavior under conditions of shock and vibration. These features make crystal resonators ideal circuit elements in high-stability narrowband applications, such as narrowband filters and highly stable oscillators. Fig. 2-24 is an equivalent circuit of a crystal resonator in the vicinity of resonance. The qualification of being in the vicinity of resonance is necessary since a quartz-crystal blank will have several resonant frequencies as determined by the crystal dimensions and the type of cut involved. These undesired resonances (or spurious responses) sometimes cause difficulties, as discussed

crystal Q

in more detail in the section on crystal filters. The Q of the crystal is defined in the standard manner for a series resonant circuit: $Q = \frac{\omega L}{R}$. The

C_0 crystal capacitance

capacitance C_0 is the sum of stray capacitance and that due to the crystal electrodes, its value is typically less than 10 pF.

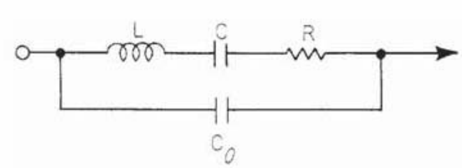


Fig. 2-24. Quartz-crystal resonator, equivalent circuit.

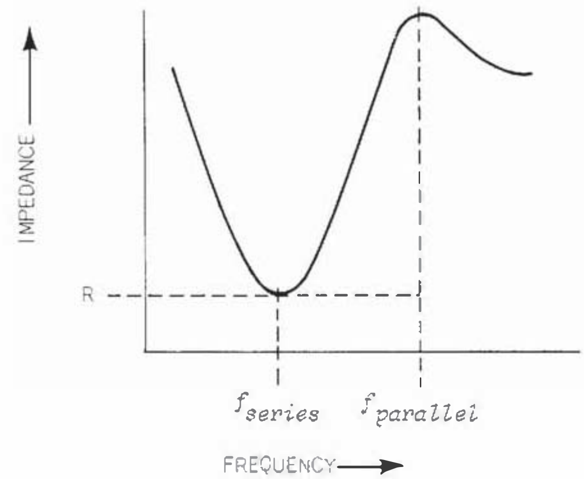


Fig. 2-25. Crystal-resonator impedance plot.

Fig. 2-25 is an impedance curve for the equivalent circuit of Fig. 2-24.

neutralizing C_0

The parallel resonance, which is due to the existance of C_0, causes undesirable effects in certain applications. For example, the signal feeding through C_0 limits the ultimate attenuation of crystal filters. As a result, it is sometimes necessary to neutralize the effect of C_0. This is accomplished by providing an additional signal path having the same amplitude-frequency characteristics as that of C_0 but with a 180° phase reversal. This secondary signal is added to the standard signal of the circuit thus cancelling the signal component which gets through by way of C_0. Capacitors C1 and C2 in the crystal discriminator of Fig. 2-8 are neutralizing capacitors, providing the secondary signal path discussed above.

REFERENCES; see page 171.

Discriminators -- B-11 pp28, 77, B-12 p987
Transmission-Line Transformers -- C-3, C-4 pp103-114
Diode Characteristics -- B-8, B-9
Phase Detectors and Samplers -- C-2
Narrow-Pulse Generators -- B-1, C-4, C-5
YIG Resonators -- C-6, C-8 Chapter 17
Quartz Crystals -- B-13 pp231-235, C-7 pp67-81

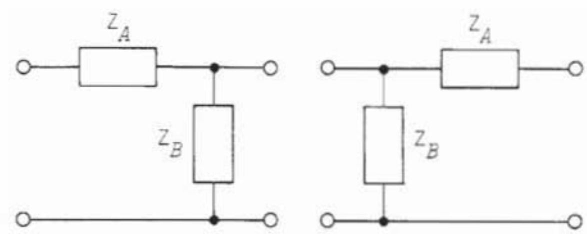


Fig. 3-1. Two half sections, when combined, form one full section.

3

FILTERS

BASIC THEORY

filter classifications

An electronic filter is a device which will selectively either pass or reject a signal having a specified band or bands of frequencies. In spite of the fairly simple function of a filter to pass or reject signals at certain frequencies, filters are classified by different systems, and designed from different points of view. For example; one can classify filters by the type of passband: low pass, high pass, bandpass, band reject; by the manner in which the filter cuts off: Tchebycheff, Butterworth; by the equivalent circuit: Constant K, m-derived; by the design procedure: image parameter; by the type of circuit elements utilized: LC, RL, transmission line, YIG; by the general performance characteristics: wideband, narrowband, microwave, variable bandwidth; by phase response: maximally linear phase response, and by many other designations. Many new filter designs have become practical through computer techniques. While the subject cannot be given a thorough treatment in this volume, the object is to present some of the basic filter theory, and then go directly to a discussion of a few of the more common types of filters used in spectrum analyzers.

sections

Most filters are made up of *sections*, where a section consists of a simple four-terminal network usually in the form of a T or a Π.

elements

half-sections

A simple T or Π section is composed of three elements, where an *element* is a single inductor or capacitor. Sections having end-to-end symmetry are usually broken up into *half-sections* as shown in Fig. 3-1.

prototype half-section: constant-K

Most filters are derived from a prototype half-section known as *constant-K*, where $Z_A \cdot Z_B = R_N^2$, where R_N is real, positive, constant. It is possible to obtain more desirable attenuation characteristics outside of the passband (higher attenuation than constant-K over a specific frequency range) by a variation known as *m-derived*, where the impedances Z_A and Z_B are modified by the factor m (m can be zero to one).

m-derived

Note also that the shunt-derived m-section has desirable impedance properties when a good match is desired to terminating resistors.

Ideally the elements of the LC filter are lossless. As a practical matter there is a certain amount of dissipation associated with these elements. We define the *dissipation factor*, δ, by:

$$\delta = \left|\frac{R}{X}\right| = \left|\frac{G}{B}\right| = \frac{1}{Q},$$

dissipation factor

where $Z = R + jx = \dfrac{1}{G + jB}$ = impedance of element, and Q is the quality factor which is equal to the ratio of energy stored to energy dissipated by the element. External loading will affect the overall Q of a device. This is particularly true in the case of distributed resonant elements, such as cavities, for instance, where the intrinsic Q of the element can be quite high. It is therefore necessary to differentiate between the Q of the unloaded element Q_U, the Q of the loaded element Q_L, and the Q of the external load Q_E. These are related by $\dfrac{1}{Q_L} = \dfrac{1}{Q_U} + \dfrac{1}{Q_E}$.

unloaded Q vs loaded Q

There are certain frequencies or frequency ranges that are normally associated with a filter. Some of these are:

cutoff frequency

Cutoff frequency - the frequency at which attenuation is increased by 3 dB; one each for low-pass or high-pass filter, and two (f_1, f_2) for a bandpass filter.

Mid-band frequency for a bandpass filter:

$$f_0 = \sqrt{f_1 f_2}$$

Bandwidth for a bandpass filter: $f_2 - f_1$

Fractional or normalized bandwidth: $w = \dfrac{f_2 - f_1}{f_0}$

filter properties and characteristics

Other terms associated with filters are:

Insertion loss refers to the loss of the filter (usually in dB) in the passband.

Ultimate attenuation refers to the maximum attainable loss outside of the passband.

Skirt selectivity refers to the rate at which attenuation increases outside of the passband.

Butterworth and Tchebycheff response

Insertion loss, ultimate attenuation and skirt selectivity are all somewhat related. Different designs provide for different compromises. Thus a *Butterworth* design provides a "maximally flat" passband, while a *Tchebycheff* design trades ripples in the passband for steeper skirts and more attenuation. This is demonstrated graphically in Figs. 3-2 and 3-3.

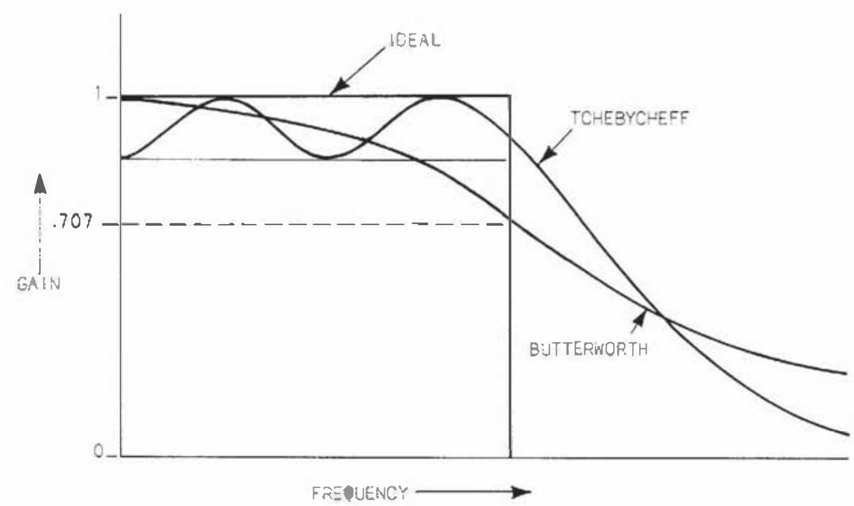


Fig. 3-2. Low-pass filter characteristics.

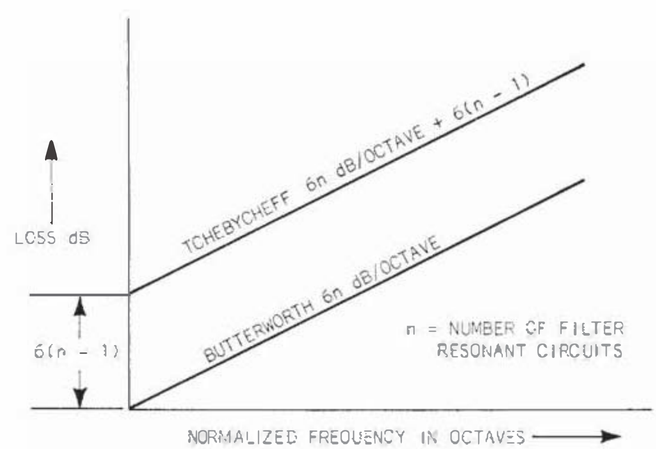


Fig. 3-3. Theoretical stop band attenuation Butterworth and Tchebycheff filters.

LC FILTERS

cascading sections

Fig. 3-4 illustrates the prototype half-sections for low-pass, high-pass, and bandpass filters. A complete filter may be constructed by cascading several prototype sections or by combining constant-K and m-derived sections. Complete low-pass filters are shown in Figs. 3-5 and 3-6.

physical design limitations

The design procedure as outlined in Fig. 3-4 is fairly simple. However, in practice, especially at the frequencies involved in many spectrum analyzer circuits, some of the element values are difficult to realize. For example, consider the problem of constructing a maximally-flat 500-MHz, 50-Ω, low-pass filter. From the basic equations we note that $L_0 = \frac{50}{6.28 \times 500 \times 10^6} = 16$ nH. It is almost impossible to build such an inductor by winding wire on a coil form. Further, the interconnections between the various elements must contribute very little inductance if the filter characteristics are to conform to the design. The solution is to use a distributed inductor printed on low-loss printed-circuit-board material, the back side of which constitutes a ground plane. The PC material must be thick enough to avoid appreciable capacitance between the printed inductor and the ground plane. The thickness of the filter can, however, not be made too great as this introduces construction difficulties as well as inductance in series with the side of the shunt capacitors that goes to ground. In addition to the above considerations it should be noted that the coil layout introduces mutual coupling between adjacent turns. In view of the above, the basic design of such a filter is approximate, the final values being determined by experiment. Fig. 3-7 is a top view of such a filter. It should be noted that in spite of its many sections (the filter contains ten capacitors) the filter is quite compact.

distributed inductor

printed-circuit filter

In general, most LC-type filters in spectrum analyzers are of conventional design and can be traced to the basic prototype circuits discussed before. The construction, however, is often unorthodox, as shown in the previous example.

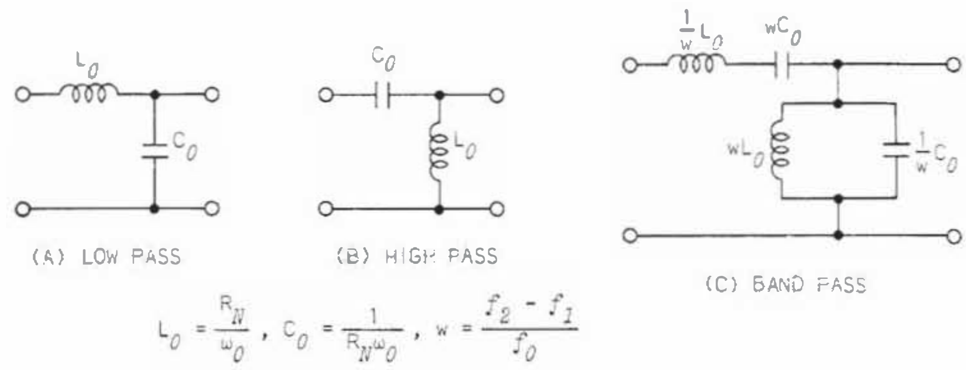


Fig. 3-4. Half-sections.

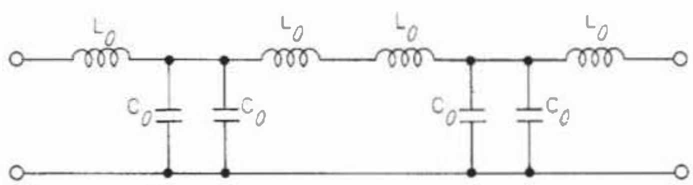


Fig. 3-5. Low-pass filter, constant-K sections.

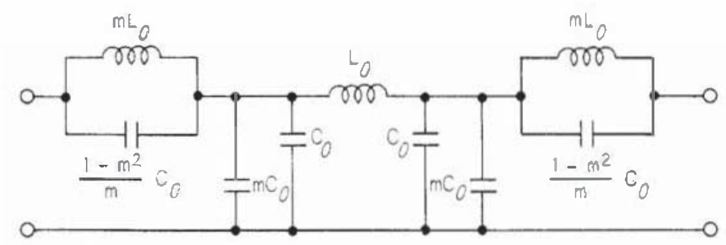


Fig. 3-6. Low-pass filter, m-derived end sections.

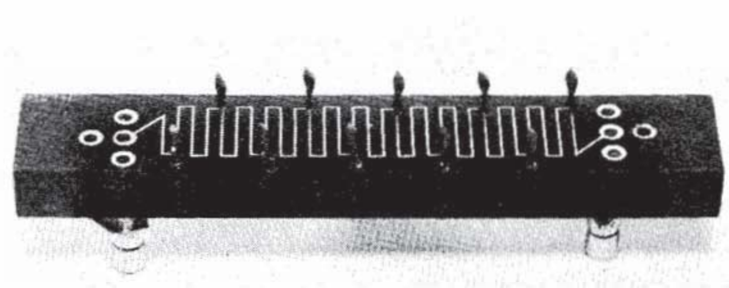


Fig. 3-7. Distributed-inductance low-pass filter.

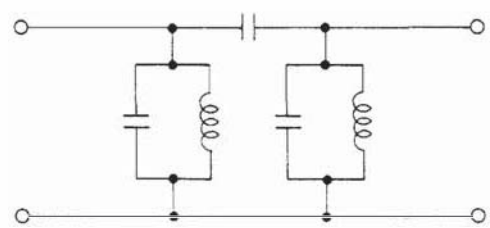


Fig. 3-8. Coupled-resonator filter.

COUPLED-RESONATOR FILTERS

There are a variety of filters which fall under the general heading of *coupled-resonator filters*. These consist essentially of an array of basic resonators, which are coupled together to form a filter. The differences between filters arise from the type of resonators: tank circuit, transmission line, cavity, etc.; coupling technique: capacitive, inductive, iris, quarter-wave transmission line, etc.; and the other filter parameters: high-pass, low-pass, Butterworth, etc. Fig. 3-8 is an example of a capacitively coupled, coupled-resonator filter design.

The filter represented schematically in Fig. 3-8 would be constructed from lumped L and C elements at lower frequencies and from distributed elements, such as transmission-line resonators, at higher frequencies. An important thing to realize is that even though these two filters would look entirely different, the basic principles governing filter behavior are the same.

As with any other type of filter, the important parameters for the coupled-resonator type are bandwidth, insertion loss, and off-band attenuation. The bandwidth is determined primarily by the impedance of the coupling element, which can be either capacitive as in Fig. 3-8, or inductive. Insertion loss is determined by the bandwidth and the inherent lossiness of the filter elements as indicated by the Q factor. Off-band attenuation rate is 6 dB per octave per resonant circuit.

bandwidth

insertion loss

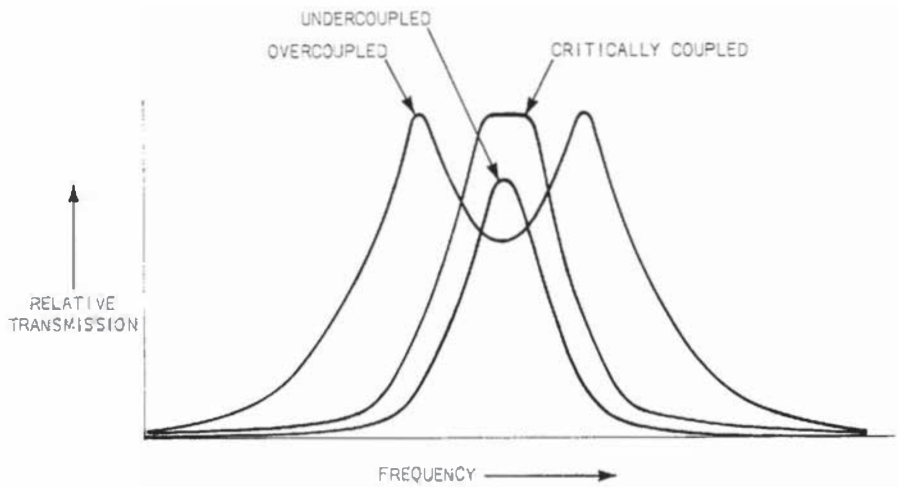


Fig. 3-9. Two coupled resonators coupling characteristics.

parameter relationship

variable resonator

resonator coupling

Under ordinary conditions only the filter designer would be interested in the relationship between the various filter parameters, e.g., between element Q and filter insertion loss. For example, the coupling capacitors in this type of filter are usually fixed, while one of the tank circuit resonators is adjustable. The effect of the variable resonator is evident since it is tuned for minimum filter loss (maximum signal output). Yet it might be of interest to note the effect of changes in interresonator coupling should repairs be necessary. This is shown in Fig. 3-9 for a two-resonator filter. The overcoupled case is best recognized by its large peaks; there are as many peaks as overcoupled resonators; while the undercoupled case is best exemplified by the increased insertion loss.

filter response

One effect controlled by the coupling resonators that is always present to some extent is asymmetry in the filter response curve. This asymmetry is most noticeable when the fractional bandwidth is relatively large (bandwidth is large when compared to center frequency). This asymmetry is reasonable in light of the frequency dependence of the coupling elements. The coupling increases with frequency for capacitive coupling elements and decreases with frequency for inductive coupling elements. Thus,

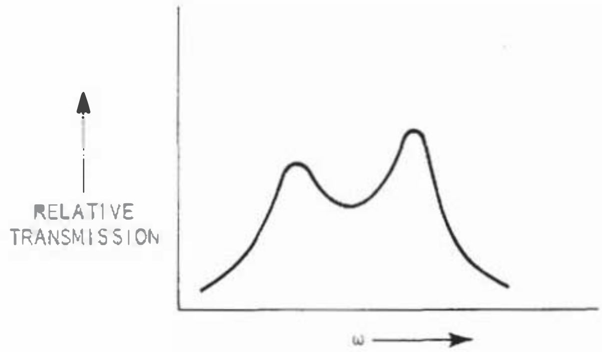


Fig. 3-10. Capacitively coupled resonators showing asymmetry effect.

asymmetry

the asymmetry with inductive-coupling elements is a mirror image of that with capacitive coupling. Actually it is possible to design the filter to be symmetrical, but it is seldom worth the trouble. Fig. 3-10 shows an asymmetrical response for a slightly overcoupled capacitively coupled filter.

RESONATOR AND COUPLING STRUCTURES

There are a variety of filters that could be considered under the heading of *coupled-resonator filters*. Most of the differences between these filters stem from construction and coupling techniques. This is particularly true at the higher frequencies where distributed resonators predominate. It is the purpose of this section to review some of the terminology and properties of the more common coupling structures.

As discussed elsewhere, a length of transmission line behaves as a resonant circuit. For example an open-circuited line will be anti-resonant (exhibit a high impedance) at the frequency where the line is half a wavelength ($\lambda/2$) long. It follows that coupled-resonator filters can be constructed from appropriately coupled transmission lines. The transmission line can be either coaxial or stripline (flat conductor printed on PC material). It has also been shown elsewhere that hollow metal cavities will behave as resonant structures, where the resonant frequency is a function of the cavity dimensions. Coupled-resonator filters can, therefore, be constructed from appropriately coupled-resonant cavities. The parameters of filters designed from distributed resonators are essentially the same as for lumped element units: bandwidth, loss, ripple, etc. There is, however, one additional parameter that must be considered, and that is higher-order passbands. These additional passbands result because distributed circuits have more than one resonant frequency; thus, an open-circuited line is anti-resonant not only at $\lambda/2$ but at all other (higher) frequencies where the line is a multiple of a half-wavelength long. A filter designed from open-circuited half-wavelength lines would have a multiplicity of spurious passbands, the first of these occurring at a frequency which is twice that of the desired passband. Very often these spurious or secondary passbands will not interfere with the intended function of the filter and so can be ignored. When this is not the case, the design must be such as to move or eliminate the undesired passband. This could take the relatively simple expedient of choosing the right kind of coupling structure (e.g., choosing between a structure that has a second passband at $2\omega_0$ and one that has a second passband at $3\omega_0$), or working out a complicated design that will tend to eliminate spurious passbands all together (e.g., combining a $2\omega_0$ structure with a $3\omega_0$ structure, or operating some resonators at $\lambda/2$ and others at λ, etc.). Mixed designs of this type will not be considered here.

properties of coupling structures

additional passbands

anti-resonant

spurious passbands

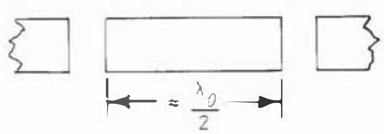


Fig. 3-11. Halfwave capacitively coupled resonators.

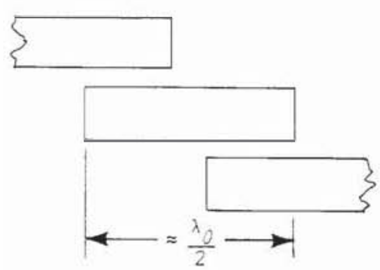


Fig. 3-12. Halfwave parallel-coupled resonators.

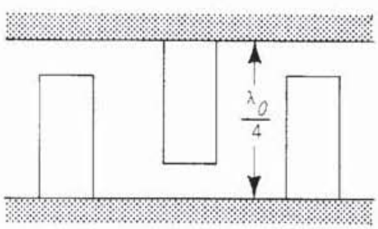


Fig. 3-13. Interdigital filter.

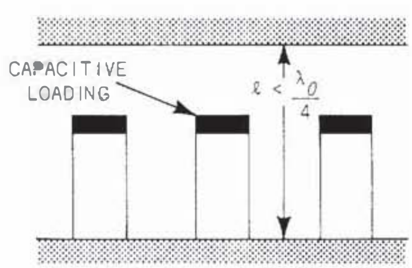


Fig. 3-14. Comb-line filter.

Figs. 3-11 thru 3-14 illustrate some common resonator and coupling structures and describe their characteristics. Symbols are defined as follows:

ω_0 = passband Center Frequency

ω_{SPB} = Center frequency of second (spurious) passband

λ_0 = wavelength at ω_0

It should be emphasized that the preceding is only a small sampling of the type of filter structures that are actually used. Further, it should be noted that there are variations in construction that may make the filter difficult to recognize as belonging to a certain class. For example, the filter shown in Fig. 3-15 is easy to recognize as being of comb-line construction, not so for that of Fig. 3-16. Here the filter is constructed of coaxial sections with iris coupling between sections. The tuning screws that provide the capacitive loading are clearly visible in the photograph. The choice of construction technique is determined by a tradeoff between the many filter parameters and ease of construction, the fit of form factor to the available space, and ultimate cost.

iris coupling

tuning screws

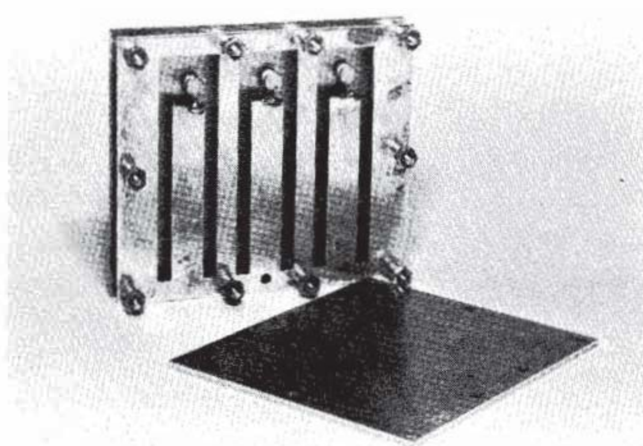


Fig. 3-15. Comb-line filter.

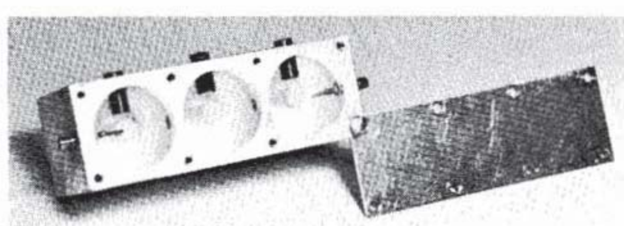


Fig. 3-16. Coaxial-structure comb-line filter.

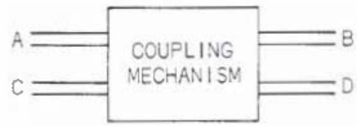


Fig. 3-17. Directional coupler.

DIRECTIONAL COUPLERS AND FILTERS

four-terminal device

A directional coupler is a four-terminal device, consisting of two transmission paths (e.g., two transmission lines) so coupled together that the power output at a port of one transmission path depends on the direction of propagation in the other transmission path. The performance of a directional coupler is determined by the *coupling factor* and the *directivity*. Referring to Fig. 3-17, these are defined as follows:

$$\text{Coupling Factor (dB)} = 10 \log_{10} \frac{P_A}{P_D}$$

coupling factor

The coupling factor specifies the amount of power delivered to a matched load at terminal D as a function of power input at terminal A, with terminals B and C terminated in matched loads.

directivity

$$\text{Directivity (dB)} = 10 \log_{10} \frac{P_D}{P_C}$$

The directivity is a measure of how well terminal C is isolated with respect to power entering at terminal A. Ideally, power entering at terminal A should not get to terminal C, only to B and D. Thus, an ideal directional coupler would have infinite directivity. In practice a good directional coupler will have a directivity of 30 to 40 dB.

Just as with filters, directional couplers can be constructed in many different ways. We shall discuss two types of couplers only: The two-hole waveguide coupler which is the simplest to understand, and the TEM-mode coupled-transmission-line coupler which is the predominant type in spectrum analyzers. Fig. 3-18 shows the basic construction of a two-hole waveguide directional coupler.

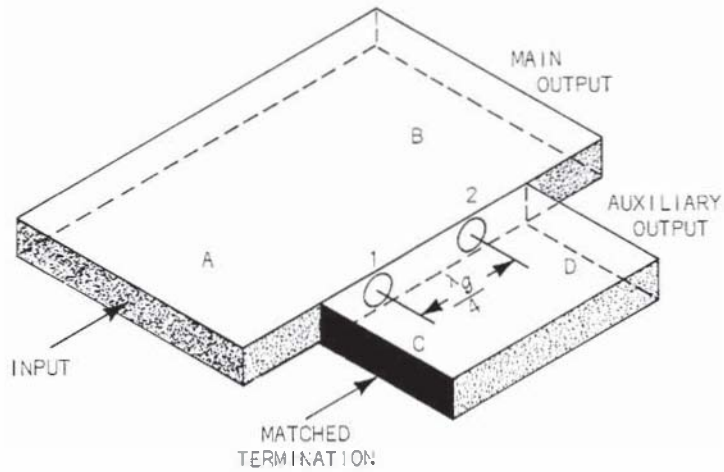


Fig. 3-18. Two-hole waveguide directional coupler.

two-hole waveguide directional coupler

Consider what happens as energy propagates from the input A to the output B. Some of this energy leaks across the wall through coupling-holes 1 and 2. The size of the coupling holes determines the amount of leakage, and thus determines the coupling factor. Now consider what happens to the energy propagating toward D and C. A certain percentage of the input energy leaks through hole 1 and propagates toward D. Some time later the input energy reaches hole 2 and some of it gets through hole 2. Since the leakage energy from hole 1 and the input energy have traveled the same distance to get from hole 1 to hole 2, the leakage from the two holes combine in phase and proceed toward D. Concurrently, some of the energy getting through hole 1 propagates toward C. Similarly, some of the energy getting through hole 2 propagates toward C. These combine in the vicinity of hole 1 where the leakage of hole 2 is half a wavelength advanced with respect to that of hole 1 (it has to travel an extra $\lambda_g/4$ on each side of the wall). Thus, the two energy waves traveling toward C are 180° out of phase and cancel. The degree of phasing and relative energy level between the two holes determines the directivity.

determine directivity

It will be noted that the two-hole coupler is a relatively narrowband device, since it depends on the fact that the distance between the coupling holes is a quarter-guide-wavelength ($\lambda_g/4$) long. In practice this type of coupler uses more than two holes thus improving the frequency characteristics. The

improve frequency charac- teristics

phasing relationship in a multi-hole coupler is naturally more difficult to analyze than that for a two-hole coupler, but the basic principle is the same. As the number of holes is increased one can conceive of the coupling becoming continuous, sort of in a line. The above is an intuitive explanation of the theory behind the TEM-mode coupled-transmission-line coupler.

coupled- transmission- line coupler

Fig. 3-19 shows the basic construction of a coupled-transmission-line coupler. It will be noted that, unlike the two-hole coupler, the coupled signal travels in a direction opposite to the input. For this reason this type of coupler is sometimes termed a *backward coupler*. This, like all other couplers, is frequency dependent and the bandwidth can be increased by cascading several sections.

Coupled-transmission-line directional couplers can be constructed out of all the materials that lend themselves to transmission-line construction. This includes wire, coaxial cable, and strip-line. Fig. 3-19 represents a stripline construction.

signal combining separating

Directional couplers are used as components in many applications. A typical application in spectrum analyzers is for signal combining or separation. A simple signal combiner-separator is the *diplexer*. A diplexer can be used in many applications, including that of separating the local-oscillator, LO, signal from the IF signal in a mixer.

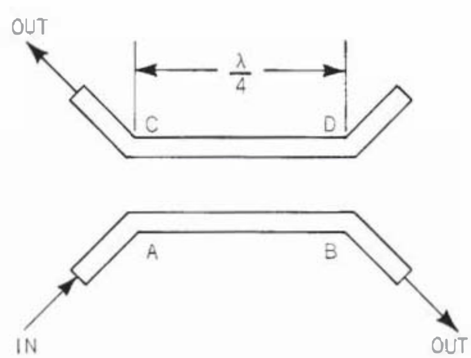


Fig. 3-19. TEM-mode coupled-transmission-line coupler.

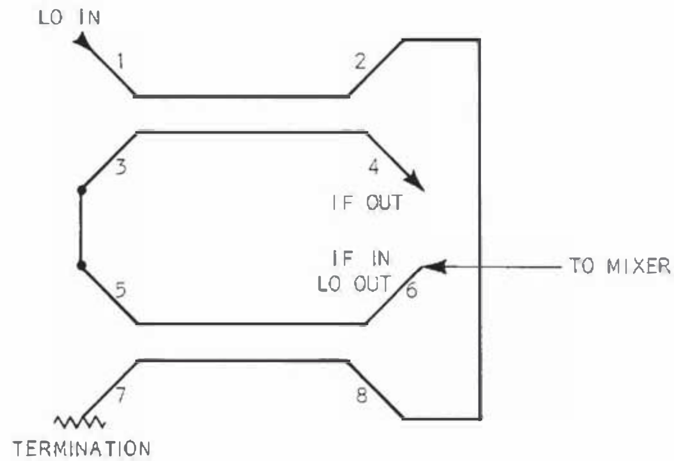


Fig. 3-20. Diplexer.

deliver to mixer

A diplexer arranged for such a purpose is shown in Fig. 3-20. The LO signal is fed into port 1; it is delivered to the mixer at port 6. The LO signal is mixed in the mixer with an incoming RF signal to produce a relatively low-frequency (e.g., the LO is 1 GHz while the IF is 0.1 GHz) IF signal. The IF signal is fed into port 6 and is delivered to the IF amplifier at port 4. The operation of the diplexer can be understood in terms of the basic properties of the directional coupler. Imagine two directional couplers that have a high degree of coupling at the local-oscillator frequency, but very little coupling at the IF frequency.

diplexer operation

The local-oscillator signal enters port 1, some of the signal goes to ports 2 and 8 (since ports 2 and 8 are connected together) and some of it goes to ports 3 and 5 (ports 3 and 5 are connected together). No local-oscillator power comes out of port 4. The local-oscillator power from port 5 splits between ports 6 and 7, likewise the power from port 8 splits between ports 6 and 7. If things are properly arranged, most of the local-oscillator power ends up in port 6. The IF signal developed in the mixer enters the diplexer at port 6 and, since there is very little coupling at this frequency, ends up at port 4.

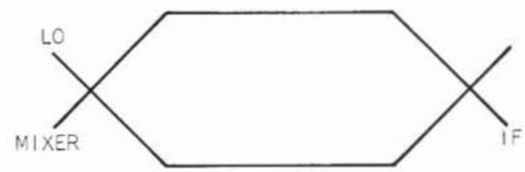


Fig. 3-21. Schematic representation of diplexer.

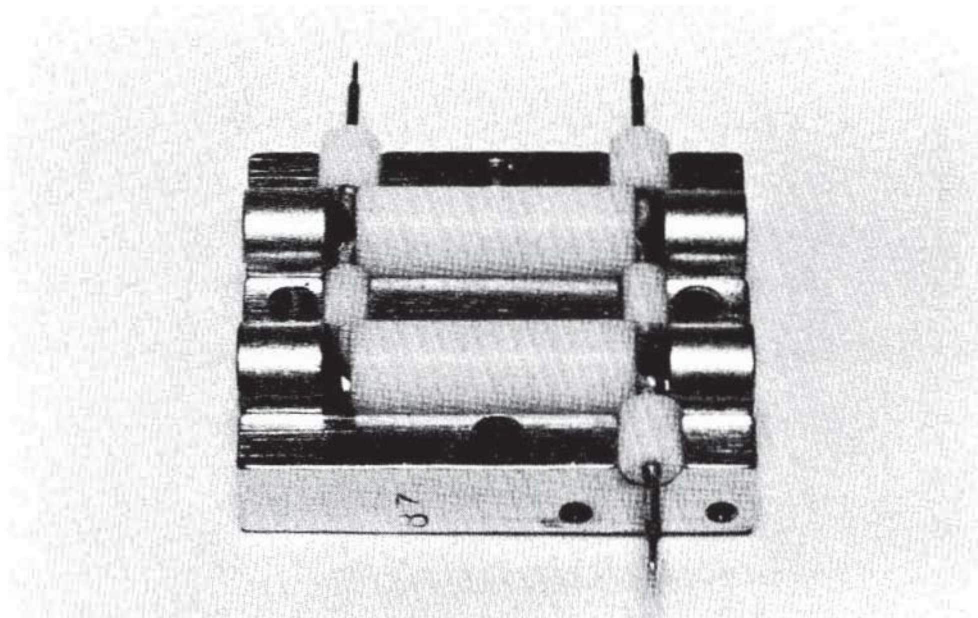


Fig. 3-22. Diplexer. Courtesy of Sage Laboratories.

Fig. 3-23. Directional filter.

The diplexer shown in Fig. 3-20 indicates coupled-transmission-line type of directional couplers. Other types of couplers can, of course, also be used. In general a directional coupler is shown as two crossed lines in schematic representation. A schematic representation of a diplexer coupler would be as shown in Fig. 3-21. Fig. 3-22 is a photograph of an actual diplexer.

directional filter

A more complex signal combiner-separator that uses directional couplers is the *directional filter.* A complete analysis of these devices, beyond the basic properties presented below, will be found in the references. Fig. 3-23 is a representation of a directional filter. With each port terminated in the characteristic impedance we get the following: Power in at 1 emerges at 4 with the frequency response of a bandpass filter, the remaining power emerges at 2 exhibiting the response of a band-reject filter for the power emerging from 4. No power emerges from 3. The unit is symmetrical so that any port can be made 1 with appropriate renumbering of the other ports. A directional filter can be thought of as being a superior type of coupler.

CRYSTAL FILTERS

Filters having very narrow percentage bandwidths (e.g., less than 1%) and high rates of cutoff (e.g., 60-dB down bandwidth, only twice that of 6-dB down) require higher-Q resonators than is practical for LC circuits. Such filters are constructed using quartz-crystal resonators -- hence the name *crystal filter.*

Subject to the limitations imposed by the nature of the resonant element, e.g., spurious resonances and need for neutralization, crystal filters can be constructed in the same configurations as other types of filters. However, the most popular configuration, at least for fixed-bandwidth filters, is the lattice construction. This configuration has self-neutralizing properties because of the symmetry, also lattice crystal filters can be built in wider bandwidths (up to about 10% is feasible) than other configurations. Since crystal filters differ from other filters only in the type of resonator used, we shall restrict the discussion of fixed-bandwidth crystal filters to the lattice configuration. Fig. 3-24 shows the standard lattice configuration. This circuit has the basic properties of a bridge as shown by the equivalent construction of Fig. 3-25. The impedances Z_A and Z_B are frequency dependent (i.e., crystals).

lattice self-neutralizing and wideband

simple lattice equivalent to bridge

A passband appears when Z_A and Z_B are equal in magnitude but of opposite polarity. That is, one is a capacitive reactance and the other an inductive reactance.

passband element

As a matter of practice, actual crystal-filter designs use the half lattice since this requires half as many crystals as the full lattice at very little sacrifice in performance. Figs. 3-26 and 3-27 are the basic and bridge configurations for the half lattice. As with other filters, lattice crystal filters can be cascaded to improve off-band rejection.

half-lattice effective

One of the major problems with crystal filters is the existence of spurious passbands. Unlike distributed circuit filters, these spurious passbands are not harmonically related to the main passband frequency. The spurious resonances in crystals are usually quite close (within several hundred kHz) to the frequency of the main response. As previously indicated these spurious resonances are connected with the crystal construction and cannot be altogether eliminated.

spurious passband

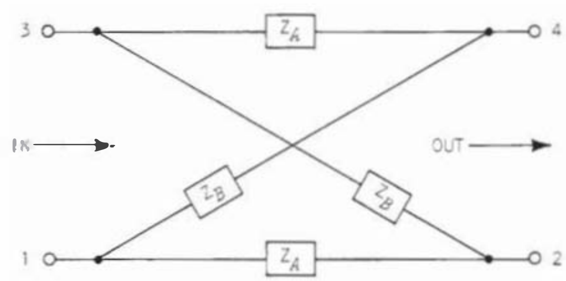


Fig. 3-24. Basic lattice.

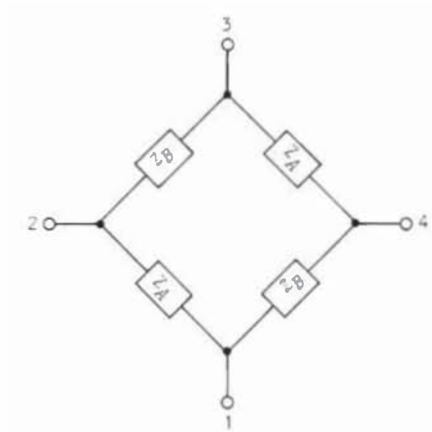


Fig. 3-25. Lattice equivalent.

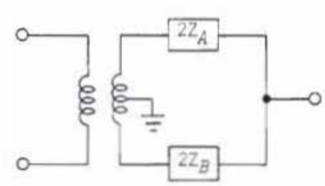


Fig. 3-26. The basic half lattice.

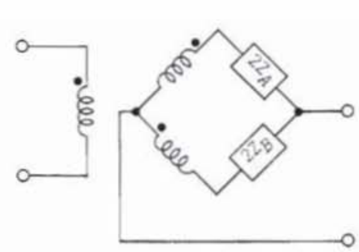


Fig. 3-27. Half-lattice bridge structure.

There are several techniques for reducing the magnitude of the spurious passbands caused by spurious crystal resonances:

avoiding spurious passbands

A) Use crystals that have no spurious resonances. This is not feasible at the present time. However, recent advances in crystal technology have led to crystals with greatly reduced spurious responses. These crystals permit construction of filters with suppression on the order of 40 dB. Figs. 3-28 and 3-29 illustrate the effect of using special crystals. In Fig. 3-28 there are three spurious passbands, one within the main passband and two outside. The largest of these is about 20-dB down. In Fig. 3-29 there are also three spurious passbands, one within the main passband and two outside. Here the use of special crystals has reduced the largest spurious response to about 40-dB down.

cascade sections

B) Use cascaded crystals or sections that have the same main response but different spurious responses. This requires a more complex and, therefore, more expensive filter, but the spurious passbands are almost completely eliminated. Figs. 3-30 and 3-31 show the response of a three-section half-lattice (6 crystals) filter, with spurious passbands suppressed to more than 60 dB. There is a very small spurious passband 1.5 divisions to the left of the main response on Fig. 3-31.

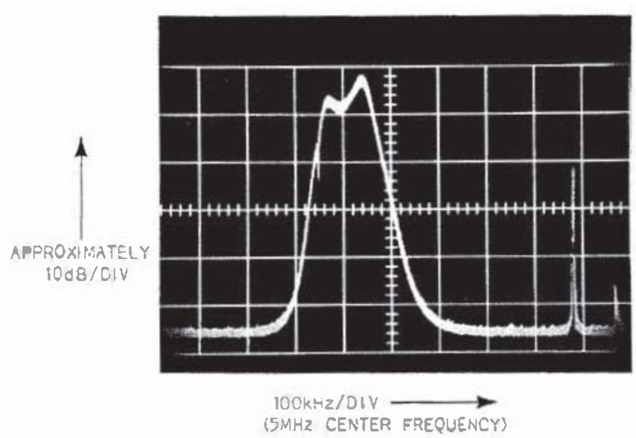


Fig. 3-28. Crystal filter, ordinary crystals, spurious passband about 20-dB down.

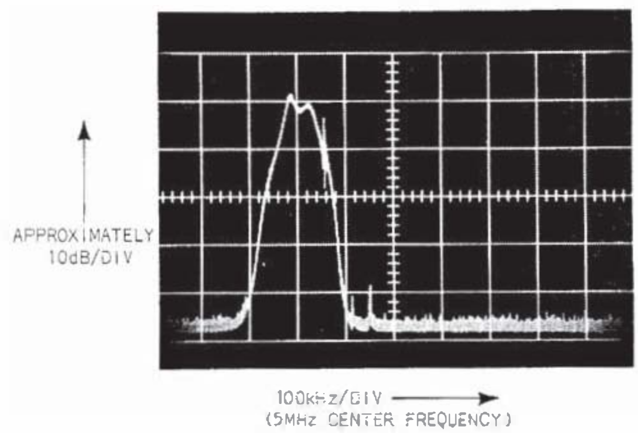


Fig. 3-29. Crystal filter, special crystals, spurious passband about 40-dB down.

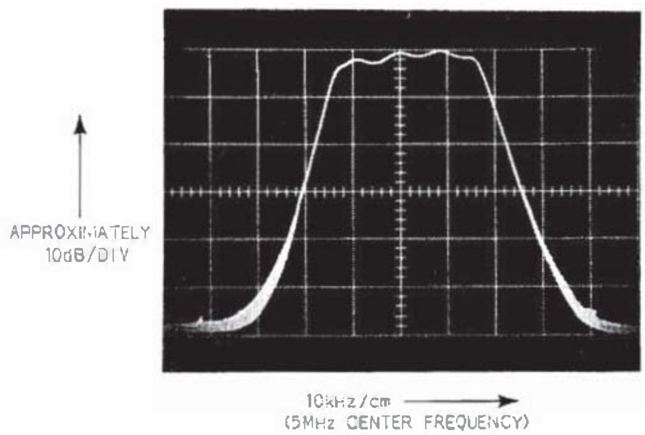


Fig. 3-30. Three-section crystal filter, fine-grain analysis.

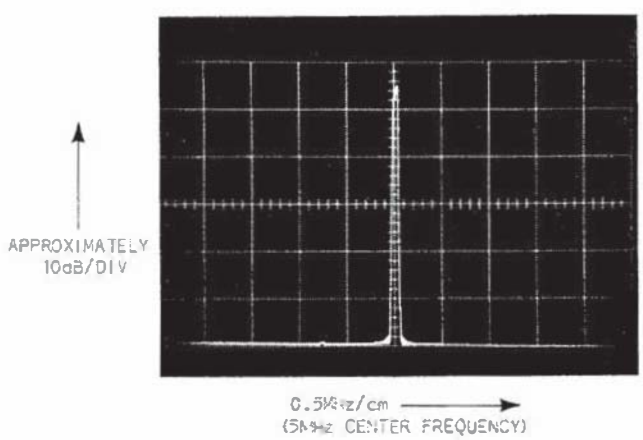


Fig. 3-31. Three-section crystal filter, wideband analysis.

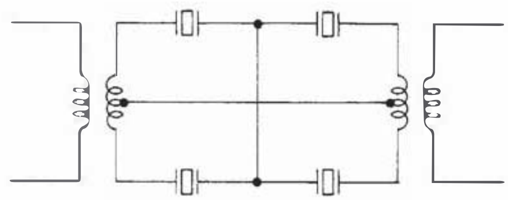


Fig. 3-32. Cascade of two half-lattice crystal-filter sections.

cascade sections

Fig. 3-32 shows how two half-lattice sections can be cascaded together. The transformers serve the dual function of supplying the needed inductance for the eliminated lattice arms as well as providing an impedance match to the load. Fig. 3-33 is a photograph showing the construction of a six-crystal (three half-lattice sections) crystal filter.

variable-resolution filters

Besides fixed-bandwidth crystal filters, spectrum analyzers also utilize variable-bandwidth crystal filters; these are commonly known as *variable-resolution* filters.

loading resistor changes Q

$\frac{V_o}{V_{in}}$

Unlike the fixed-bandwidth filters, variable-bandwidth crystal filters are mainly of ladder construction, where the bandwidth variation is accomplished by adjustment of the circuit Q. An elementary filter of this type is shown in Fig. 3-34. The crystal has been represented by the ideal equivalent circuit (a series-resonant LC) where the impedance goes to a low value at the resonant frequency, f_0. Similarly for the parallel circuit the impedance is a maximum at f_0, and the bandwidth is almost completely determined by the loading resistor R, assuming that the next circuit is sufficiently tapped down on the coil to provide negligible loading. As for any series-parallel circuit, the output V_o is related to the input V_{in} by the simple voltage division between the series-impedance Z_A and the parallel-impedance Z_B --

$$V_o = V_{in} \frac{Z_B}{Z_B + Z_A}.$$

→ approaches

From the above it is clear that $V_o \to V_{in}$ as $Z_A \to 0$, $Z_B \to \infty$, and $V_o \to 0$ as $Z_A \to \infty$, $Z_B \to 0$; and V_o takes on values between zero and V_{in} depending on the impedance relationship between Z_A and Z_B; for example, at $Z_A = Z_B$, $V_o = 1/2\ V_{in}$. Since Z_A and Z_B

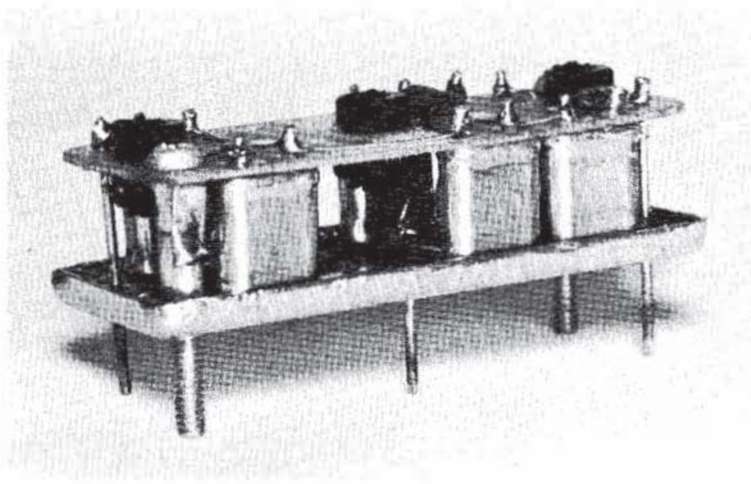


Fig. 3-33. Three-section crystal filter.
Courtesy of CF Networks, Inc.

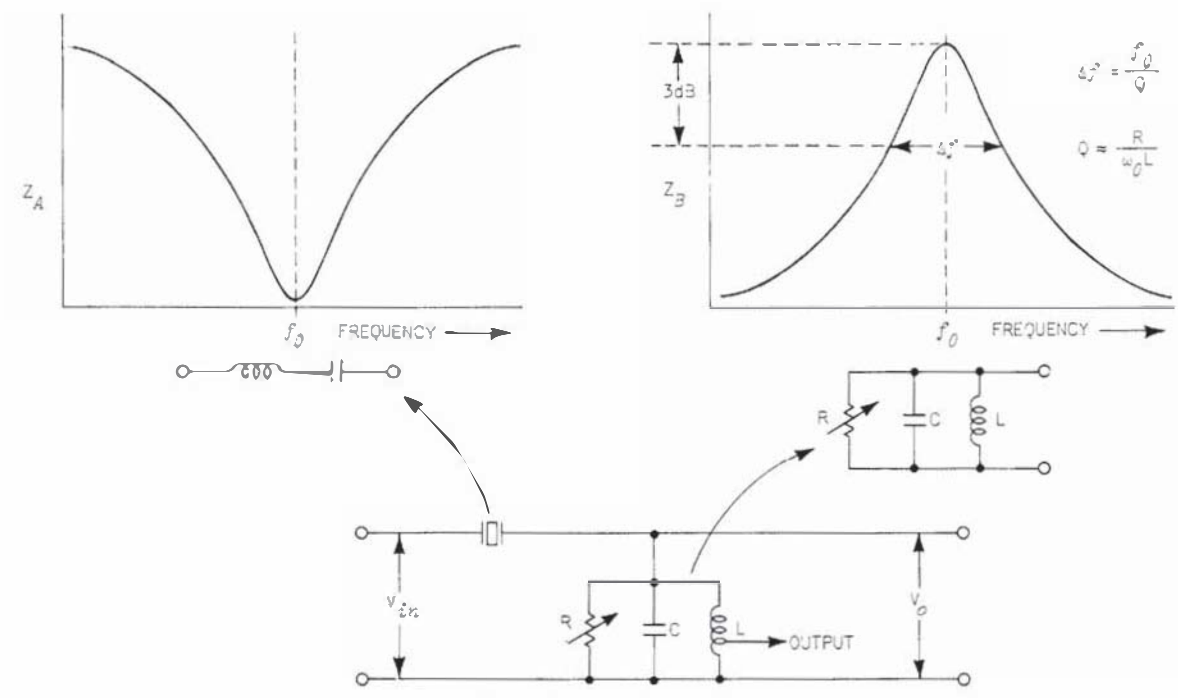


Fig. 3-34. Variable-bandwidth crystal filter
and impedance response curves.

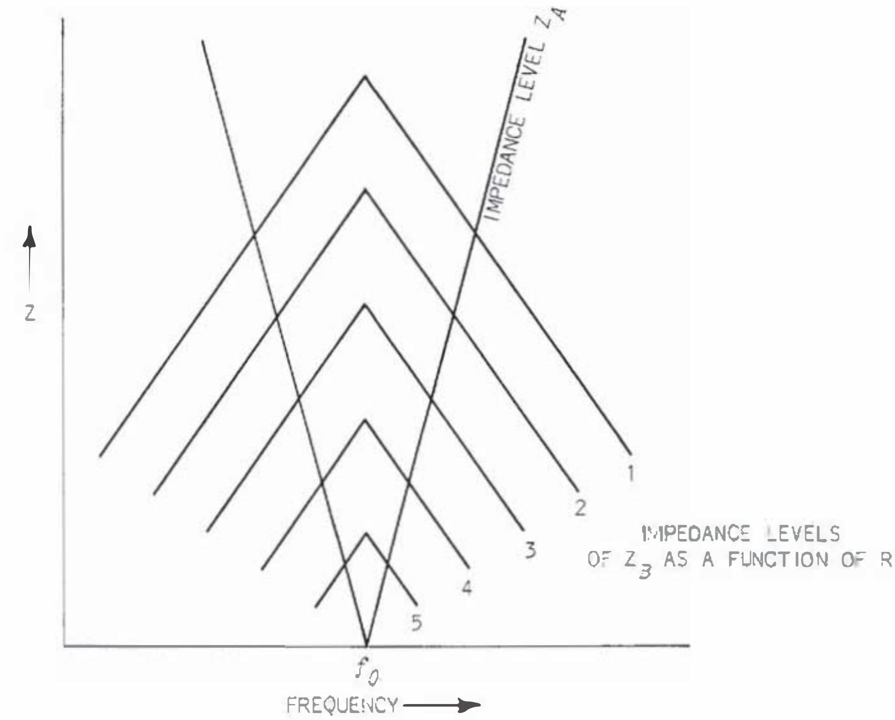


Fig. 3-35. Impedance relationships in variable-bandwidth crystal filter.

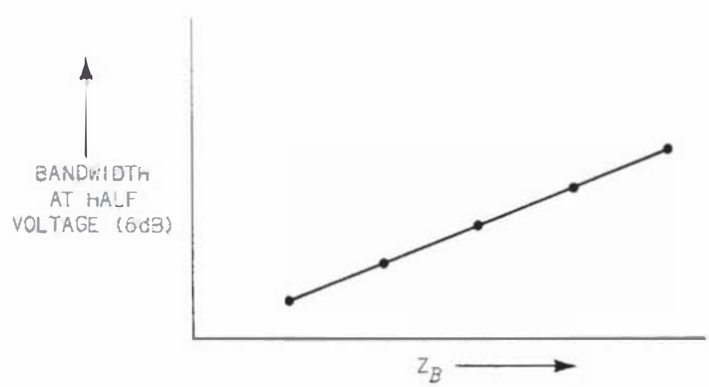


Fig. 3-36. 6-dB bandwidth as a function of shunt impedance Z_B.

are functions of frequency it follows that, for a constant V_{in}, V_o will be a function of frequency. Further, the relationship between V_{in} and V_o, which is a function of frequency, will change as the value of Z_B is changed by adjustment of R. In other words, V_o is a function of frequency: The network is a filter; the relationship between V_o and V_{in} is changed by adjustment of R: The network is an adjustable filter.

To illustrate the fact that the circuit of Fig. 3-34 is that of a variable-bandwidth filter, consider the idealized graphical construction in Fig. 3-35.

response relationship

In Fig. 3-35 the resonance curves of Z_A (a crystal) and Z_B a high-Q LC tank circuit are idealized as triangular responses. The level of Z_A is fixed while that of Z_B is adjustable. At the center frequency, f_0, Z_A is very small and Z_B very large so that V_o is essentially equal to V_{in}. At some frequency on either side of f_0, $Z_B = Z_A$ and $V_o = 1/2\ V_{in}$. This is the half-voltage or 6-dB bandwidth plotted in Fig. 3-36 as a function of Z_B. In practice the relationship between the bandwidth and Z_B is not linear as shown by 3-36, since the actual resonance curves do not have straight line sides as idealized in 3-35.

neutral-ization for off-band attenuation

In actual usage the circuit of 3-34 is modified to provide crystal shunt-capacity neutralization and to compensate for changes in output level as a function of bandwidth. The shunt-capacity neutralization is necessary in order to obtain reasonable (more than 30 dB) off-band attenuation. The amplitude compensation takes care of loss in output amplitude at very narrow bandwidths where the tank-circuit loading becomes comparable to the crystal impedance at resonance (about 20 to 50 Ω). For example, the theoretical loss in amplitude would be 6 dB when the loading impedance is reduced to the level of the crystal at resonance. Other circuit complications have to do with temperature compensation for the bandwidth, or gain; or gain as a function of bandwidth.

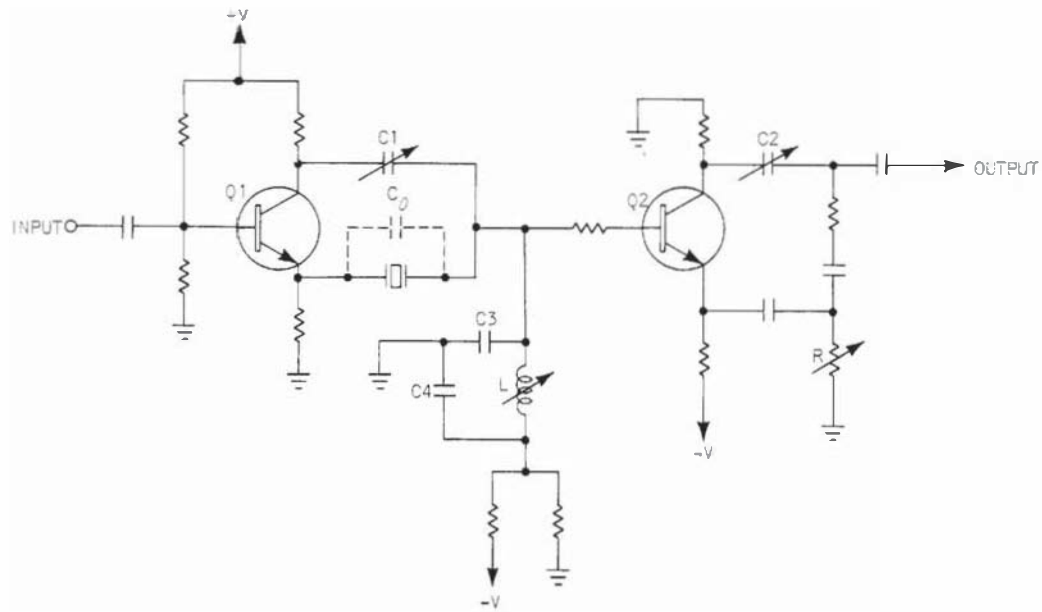


Fig. 3-37. One stage of a variable-bandwidth filter.

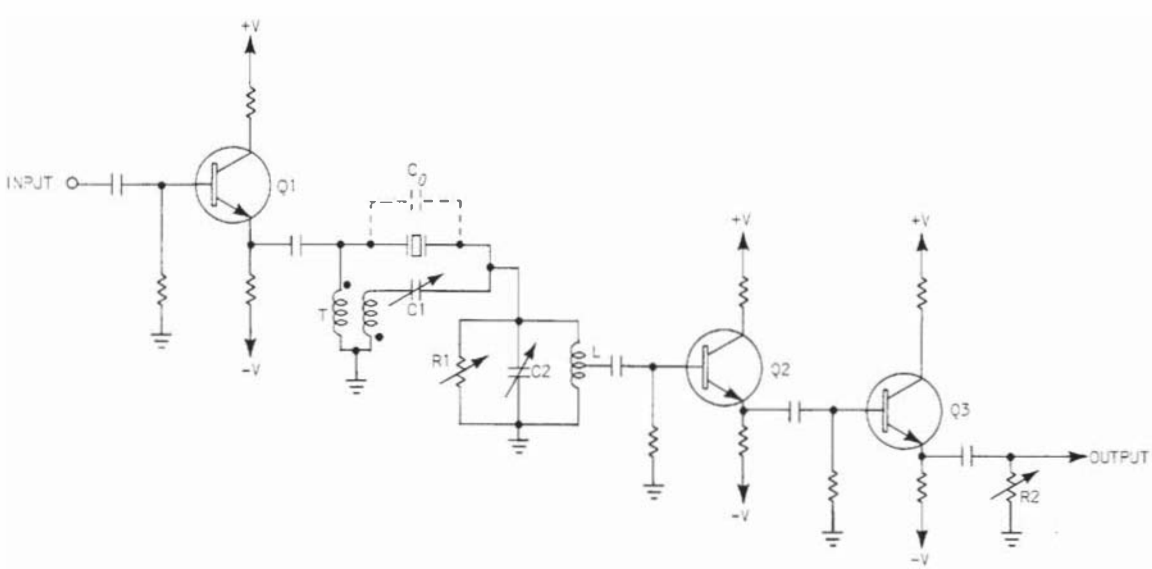


Fig. 3-38. Variable-bandwidth filter with transformer-coupled neutralization.

practical circuits

amplitude compensation versus bandwidth

One practical variable-bandwidth filter configuration is shown in Fig. 3-37. The crystal is driven by phase splitter, Q1, so that C1 can be adjusted to compensate (neutralize) for the effects of the crystal shunt-capacity C_0. The high-impedance load is provided by the tank circuit formed by capacitors C3 and C4 and inductor L. Bandwidth control is by means of variable resistor R, which controls the emitter load and hence the input impedance of amplifier Q2. Amplitude compensation as a function of bandwidth, which is a function of R setting is provided by gain adjustment C2. As with any other type of filter, several sections of this filter can be cascaded.

neutralizing

precise amplitude control

A different variation of the basic circuit of 3-34 is that of Fig. 3-38. Here the crystal is driven by the emitter-follower Q1. Transformer T provides the necessary 180° phase reversal for the neutralizing capacitor C1. The signal drives the high-impedance tank circuit formed by C2 and L. R1 controls the bandwidth by controlling tank-circuit Q. Emitter-follower Q2 provides isolation between the tank circuit and the amplitude-equalizing circuit of R2 in the emitter of Q3. This circuit lends itself to precise amplitude control since R2 can be connected to the same front-panel control as R1.

It should be clear from these examples that though many different variable-bandwidth crystal-filter configurations are possible, most of these can be explained in terms of the basic circuit of Fig. 3-34.

REFERENCES; see page 171.

Filters -- B-4 pp115-127, B-5 Chapters 9 and 10, B-12 Chapters 26 and 27, B-13 Chapters 6, 7, and 8, C-8, C-14
Crystal Filters -- B-13 pp231-235, C-7 pp67-81
Directional Couplers -- B-7 Chapter 14, C-8

4

AMPLIFIERS

amplifier-function criteria

Spectrum-analyzer amplifiers can be classified in different ways: Narrowband - wideband; high gain - low gain; low noise - noise doesn't matter; front end - output; linear - logarithmic; and a combination of these. Because of the manner in which spectrum analyzers are constructed, it is found that many of these characteristics tend to go together. Thus, wideband amplifiers are generally linear low-noise front-end amplifiers, while logarithmic amplifiers are generally high-gain output amplifiers. In the following sections we shall consider the special characteristics of narrowband, wideband, low-noise, high-gain, and logarithmic amplifiers.

NOISE CHARACTERISTICS OF AMPLIFIERS

noise figure

An ideal amplifier would increase the level of the input signal without introducing any additional noise, but actual amplifiers do contribute noise. The noisiness of amplifiers is defined by a figure of merit called *noise figure,* (F). Noise figure is defined as

$$F = \frac{\text{signal to noise ratio in ideal case}}{\text{signal to noise ratio in actual case}} = \frac{S/N \text{ (ideal)}}{S/N \text{ (actual)}}$$

Since the actual S/N cannot be greater than the ideal S/N (the actual noise is greater or equal to that of the ideal case), it follows that noise-figure numbers range from unity, for an ideal amplifier, and upward. Sometimes noise figure is given in decibels, thus, $F_{dB} = 10 \log_{10} F$. F_{dB} goes from zero, for an ideal amplifier, and up.

resistance-equivalent noise

It should be noted that an ideal amplifier is not completely noise free, but rather contributes the same amount of noise as a resistor whose resistance is equal to the input impedance (assuming a resistive input impedance) of the amplifier. The noise power that a resistor can deliver into a matched load is given by

noise power

$N = kTB$, where k = Boltzmann's constant = 1.37×10^{-23} watt-seconds/deg
T = absolute temperature (i.e. 273° + degrees centigrade)
B = the noise bandwidth.

This noise, N, is variously known as Johnson noise; J.B. Johnson reported this effect in 1928; and thermal noise: Noise due to temperature effects. It is clear from the above that we must specify an ambient temperature for our hypothetical ideal amplifier. This temperature has been standardized at T = 290° abs = 17°C = 63°F. Using the above values of k and T we can calculate that $N = 4 \times 10^{-21}$ watts/cycle or $N = 4 \times 10^{-12}$ mW/MHz which is equal to -114 dBm/MHz. The sensitivity of a 1-MHz-wide ideal amplifier is 114 dB below one milliwatt*. In practice this number will be degraded by the noise figure of the amplifier. Amplifiers are usually made up of several stages of gain, each stage contributing its own noise to the overall system. It is intuitively logical that the noise of the first stage should count the most since that noise is amplified the most. In a cascade of amplifier stages a, b, and c, having noise figures F_a, F_b, and F_c, and overall noise figure F_{abc}, and power gains G_a, G_b, and G_c -- the relationship between parameters is:

temperature

-114 dBm/MHz

$$F_{abc} = F_a + \frac{F_b - 1}{G_a} + \frac{F_c - 1}{G_a G_b}$$

From this it is clear that a low-noise-figure high-gain stage can be followed by a poorer-noise-figure stage than a low-noise-figure low-gain stage.

*This can be improved by cooling the amplifier since N is proportional to T.

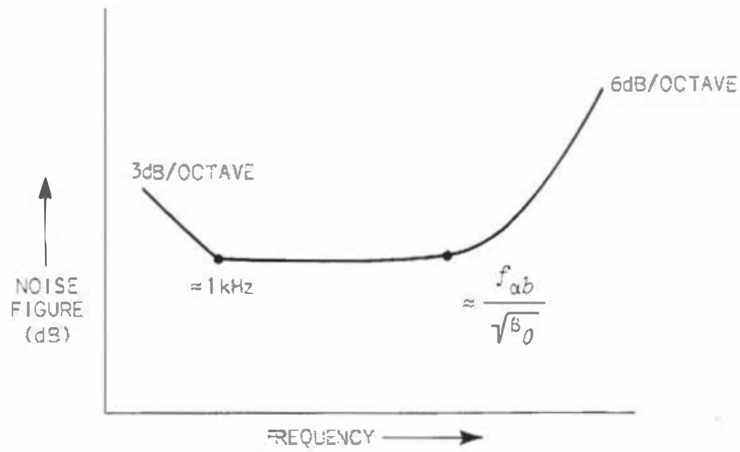


Fig. 4-1. Transistor-noise spectrum.

transistor noise

We have now reached the stage where it is logical to consider the noise properties of the gain producing element, in our case a transistor. Fig. 4-1 shows the basic transistor-noise spectrum.

flicker, $1/f$ noise

At some relatively low frequency, the noise begins to increase with decreasing frequency at about 3 dB per octave. This is referred to as *flicker noise* or *one-over-f noise*.

shot noise

Shot noise predominates at higher frequencies, and the noise figure is relatively constant as frequency increases to an approximate point:

$$f_N = \frac{f_{\alpha b}}{\sqrt{\beta_0}}$$

where,

$f_{\alpha b}$ = alpha cutoff frequency

β_0 = low-frequency current gain.

Here the noise figure begins to increase at about 6 dB per octave.

In addition to these frequency effects, transistor noise is also dependent on emitter current and source resistance. The optimum (lowest noise figure) source resistance will vary somewhat from transistor to transistor but is generally in the vicinity of 1000 Ω in the 1-MHz region and 200 Ω in the 100-MHz region. Most small-signal amplifier transistors have optimum noise figure at relatively low emitter currents (below 0.5 mA). However, there is now

available a class of high-power low-noise transistors that do not degrade until substantial currents are reached (e.g., 4-dB noise figure at 50 mA). Even though the numbers may vary from transistor to transistor there is a particular range of emitter currents for each where noise figure is best.

noise vs transistor parameters

The following rules are important in the design of low-noise transistor amplifiers: 1) Use a low-noise transistor. A properly optimized low-noise transistor has an ultimate noise figure less than 4 dB. Some recently introduced transistors are supposed to have noise figures as low as 1 dB. 2) Reduce flicker noise by using as high a low-frequency cutoff as possible. 3) Design the bias circuit for an appropriate emitter current, and provide the optimum source resistance. If necessary use a transformer to generate the appropriate source resistance. 4) Avoid the 6-dB-per-octave rise in noise figure by using a high-frequency transistor. Note that for an alpha-cutoff of 100 MHz and a β_0 of 50, the noise figure starts to rise at only 14.1 MHz.

noise figure design

These conditions explain why one might frequently find an amplifier composed of a cascade of stages identical except for the first stage. Here the transistor, the coupling and the bias arrangement might be modified in order to insure good noise characteristics.

noise-critical first stage

HIGH-GAIN, LINEAR AMPLIFIERS

Theoretically one can get as much gain as desired simply by cascading a sufficient number of amplifier stages. In practice high-gain amplifiers are difficult to build because of shielding and oscillation problems. We shall not concern ourselves here with these circuit-layout problems. A high-gain-amplifier problem that is of interest here is that of linearity. It is difficult to design high-gain amplifiers that are also highly linear; the higher the amplifier gain the poorer the linearity tends to be, this results because for equal input levels, high-gain amplifiers have to handle a wider range of signal amplitudes than low-gain amplifiers.

maintain linearity

There are two distinct aspects of spectrum-analyzer linearity that must be considered. One is a nonlinear input-output transfer characteristic: A decrease in the accuracy of amplitude calibration. Amplitude calibration error is usually specified to a few percent, typically 3%. Unless the amplifier nonlinearity is appreciable (e.g., 1%), it should have minor effect on amplitude calibration accuracy.

signal distortion

The other effect of amplifier nonlinearity is signal distortion, which can be much more serious than the calibration accuracy problem, since fairly low levels of nonlinearity can cause appreciable distortion. There are several different types of distortion, the most serious being the introduction of odd-order intermodulation (IM) products.

nonlinear circuit a mixer

third-order products

Intermodulation, as the name implies, refers to the interaction or modulation of one signal by another. For intermodulation distortion to occur it is necessary that more than one signal be present. Since spectrum analyzers are rarely used to analyze signals having only a single frequency component, IM is of much importance in spectrum analysis. Consider two signal components, one at a frequency f, the other at a slightly higher frequency $f + \Delta f$. The nonlinear amplifier can be considered as being a mixer where either signal can be considered as the local oscillator. As with any other mixer the two signals and their harmonics combine to generate sum and difference frequencies. Thus we have $f + (f + \Delta f) = 2f + \Delta f$, and $(f + \Delta f) - f = \Delta f$ as the second-order mixer products where second order denotes the sum of the harmonic numbers of f and $(f + \Delta f)$. Similarly for the third-order products we have $2f + (f + \Delta f) = 3f + \Delta f$, $2(f + \Delta f) + f = 3f + 2\Delta f$, $2(f + \Delta f) - f = f + 2\Delta f$, $2f - (f + \Delta f) = f - \Delta f$. It will be observed that the frequencies of the second-order products are relatively far removed from the original frequencies, but for the third-order products there are two new signals ($f + 2\Delta f$, $f - \Delta f$) whose frequencies are quite close (within Δf) to those of the original signals. These close-in signals will appear on the spectrum-analyzer readout and limit the capabilities of the instrument. A complete analysis shows that odd-order products are harmful while even-order products are not.

The degree of intermodulation is dependent both on the amplifier linearity and the signal levels, increasing in a nonlinear fashion with both. It follows that an amplifier handling low-level signals can be permitted to be less linear than one handling high-level signals. Further, the distortion falls off with increasing product orders. Fifth-order distortion is smaller than fourth-order distortion. Since distortion is a function of gain and linearity, the choice of proper bias point is very important. Sometimes an amplifier will appear to be biased in a nonoptimal or generally unusual manner. Such bias arrangements are frequently dictated by IM-distortion considerations.

bias point versus IM distortion

NARROWBAND AMPLIFIERS

narrowband

There is no one specific bandwidth that one can state as being the demarcation line between wide and narrowband amplifiers. In general narrowband amplifiers in spectrum analyzers have fractional bandwidths that are less than 10%. Most of the amplifiers in spectrum analyzers are narrowband; the major exception being the first amplifier in a swept-IF type of system.

interstage coupling versus selectivity

bandpass control

Most spectrum-analyzer linear narrowband amplifiers are straightforward in design. Interstage coupling is essentially of two types. In one case the interstage coupling provides all of the amplifier selectivity, while in the other case selectivity is determined by input or output filters while interstage coupling is relatively broadband. The use of broadband coupling structures and external filters permits easy interstage impedance optimization and precise bandpass control. The cost is the inclusion of external filters. Broadband coupling structures are usually of the single-tuned variety demonstrated in Fig. 4-2. Besides differences in coupling, it will be found that some amplifiers use neutralization while others do not. High-gain grounded-emitter amplifiers have a tendency to oscillate unless neutralized. Neutralization is avoided by the use of transistors having low base-to-collector capacitance and by operating each stage at reduced gain. Fig. 4-3 shows a transformer-coupled amplifier stage including an adjustable neutralizing-capacitor C_n.

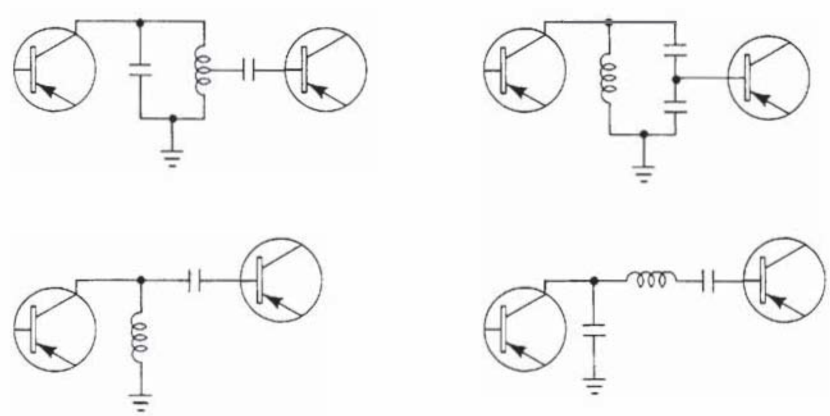


Fig. 4-2. Narrowband-amplifier single-tuned coupling structures.

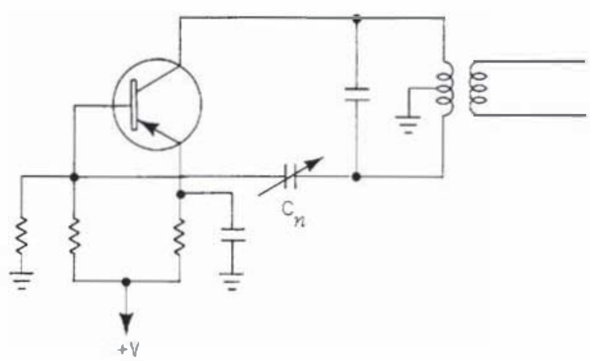


Fig. 4-3. Transformer-coupled amplifier with neutralization.

WIDEBAND AMPLIFIERS

front-end critical

Though most of the gain in a spectrum analyzer is provided by narrowband amplifiers, the characteristics of the front-end wideband amplifier in a swept-IF type of system is of the utmost importance. The amplifier must have very wide bandwidth, typically over 50% fractional bandwidth, a low-noise figure (the noise figure subtracts directly from the sensitivity of the analyzer), a flat-gain characteristic over the passband, better than ±0.25 dB, and good skirt selectivity.

Although the above requirements are not contradictory, it is nevertheless difficult to optimize all the parameters. One technique that makes both the design and subsequent circuit alignment easier is to separate the functions so that one circuit or section can be adjusted to optimize one rather than several functions. In

wideband amplifiers this indicates a design where skirt selectivity is determined by external filters so that interstage coupling can be optimized for noise, or flatness, or some other parameter without worrying about bandpass control. This technique has been used in Tektronix spectrum analyzers but with the modification of providing some minimal amount of selectivity in interstage coupling. The interstage selectivity serves the purpose of limiting the IF-generated noise power, and so prevents sensitivity degradation due to noise intermodulation: Noise power at the output of an amplifier increases not only with increasing gain but with increasing bandwidth. If noise of sufficient amplitude is present, noise components will intermodulate creating new noise components at frequencies not originally present. Some of these new noise components will fall within the passband of the narrowband amplifier following the wideband amplifier, thus increasing the total system noise output and degrading sensitivity. Therefore, it is good practice to make the wideband-amplifier gain no greater than necessary and to provide some minimal interstage bandwidth restriction even when external filters are used.

noise power versus gain and BW

limit wideband gain and interstage BW

There are essentially two basic methods for obtaining wideband amplifiers: Distributed-amplifier techniques and feedback techniques. Distributed amplifiers are not used in spectrum analyzers, so we shall consider feedback methods only.

Most wideband spectrum-analyzer amplifier designs use either fixed or adjustable frequency-selective feedback, or peaking, to provide a flat frequency response over the range of interest. The peaking is adjusted for the best response in the passband; out-of-band attenuation is of minor interest since this is provided for by separate filters. Two basic types of peaking can be used: Emitter feedback as shown in Fig. 4-4, collector feedback as shown in Fig. 4-5, or various combinations of both as indicated by Fig. 4-6. Many variations are possible, including emitter and collector peaking of the same amplifier stage as shown in Fig. 4-7. Here *C* forms the emitter peaking, *L* provides collector peaking, and *T* is a transformer for coupling to the next stage. Amplifiers of the type shown in Fig. 4-7 have particularly desirable flatness characteristics: Better than ±0.25 dB for 100-MHz bandwidth at 200-MHz center frequency.

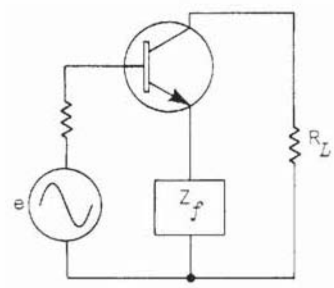


Fig. 4-4. Emitter feedback.

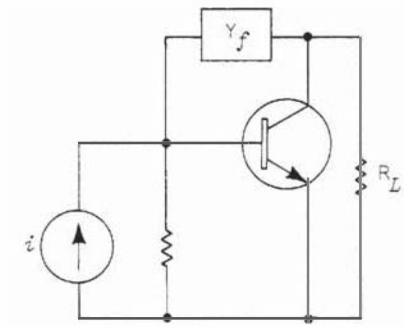


Fig. 4-5. Collector feedback.

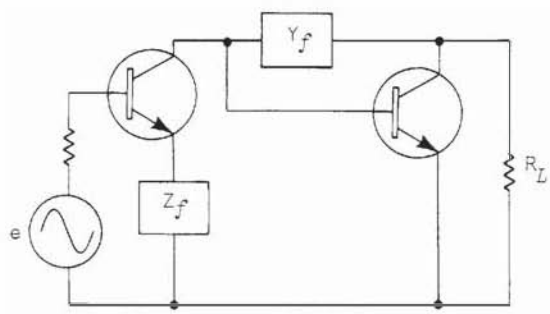


Fig. 4-6. Combination emitter-collector feedback.

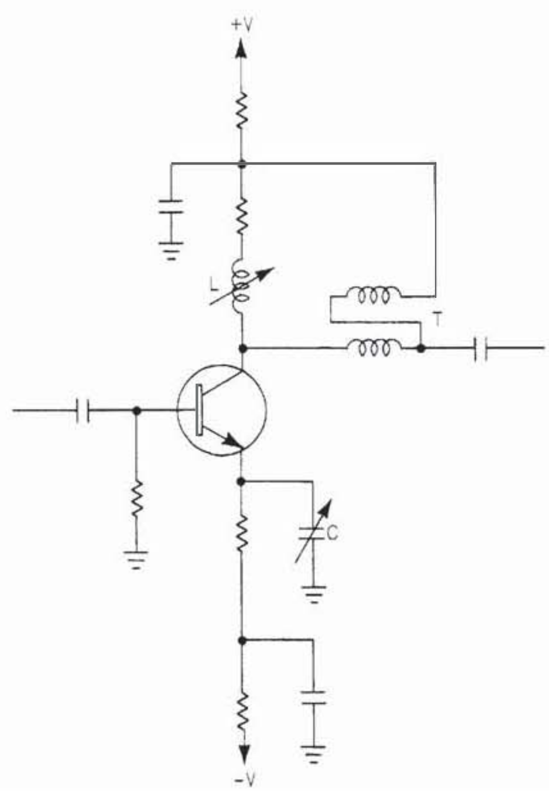


Fig. 4-7. Wideband amplifier stage, detailed circuit.

LOGARITHMIC AMPLIFIERS

extend dynamic range

Logarithmic amplifiers (log-amps) serve to extend the dynamic range of spectrum analyzers by a logarithmic input-output function: $V_{out} = k \log_x V_{in}$ (where the base x is determined by design requirements); V_{out} changes little while V_{in} changes much. One means of providing the logarithmic relationship is to use a logarithmic detector, constructed from a properly biased diode matrix, and driven by a linear amplifier. The combination of a linear amplifier in cascade with a logarithmic detector, which has increasing loss as a function of input level, produces a logarithmic-gain function. Another technique is to design an amplifier that has high gain at low input levels and low gain at high input levels, thus producing a logarithmic input-output relationship. Both types of log-amp circuits are used in spectrum analyzers.

$V_{out} =$ $k \log_x V_{in}$

linear amplifier, log detector

The log-detector linear-amplifier combination is easy to design, fabricate, and adjust. This system is relatively inexpensive and will work at all IF frequencies where linear detectors are available. The true logarithmic amplifier, on the other hand, has superior performance and provides a highly accurate, calibratable, logarithmic function and can provide a greater dynamic range than a detector system. Until recently most spectrum-analyzer log-amp systems were of the detector type, primarily due to the cost of constructing true IF log-amps and the lack of components for systems above the MHz area. However, it is now possible to construct true IF log-amps operating above 10 MHz. The trend appears to be away from the logging detector toward the true-logarithmic amplifier.

log detector less costly

The linear amplifier of a log-detector linear-amplifier system was considered elsewhere. We shall, therefore, consider only the log-detector. Fig. 4-8 is a schematic of a logarithmic detector. Diodes D1 and D2 form a standard linear detector, the logarithmic characteristic will therefore take place at low frequencies. At low signal levels diodes D3 and D4 present a relatively high

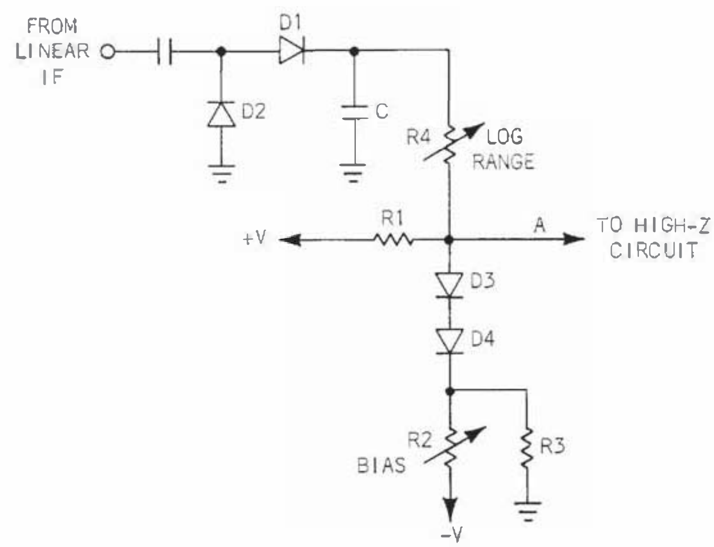


Fig. 4-8. Logarithmic detector.

impedance so that very little current flows through R4. The result is that low signal levels appear at point A unattenuated. At higher signal levels D3 and D4 turn-on, current flows through R4, and the output level at A is reduced. A logarithmic volt-ampere characteristic of D3 and D4 provides a logarithmic transfer function between the detector input and point A. Resistor R4 controls the input range over which the logarithmic function takes place, while R1, R2, R3 set the bias for the log-function diodes to control the logarithmic function. Various other log-detector circuits have been used, however they all work on the same principle.

linear-limiting log amplifier

There are many diverse logarithmic-amplifier design concepts. In general these can be separated into two different types. One is composed of a cascade of amplifier stages each of which exhibits a logarithmic gain function. This logarithmic gain function is obtained by the use of nonlinear (diode) feedback or interstage coupling. This type of amplifier has been supplanted by a more powerful technique based on an array of linear amplifier stages combined with controlled limiting. We shall now discuss the linear-limiting type of log-amp in more detail.

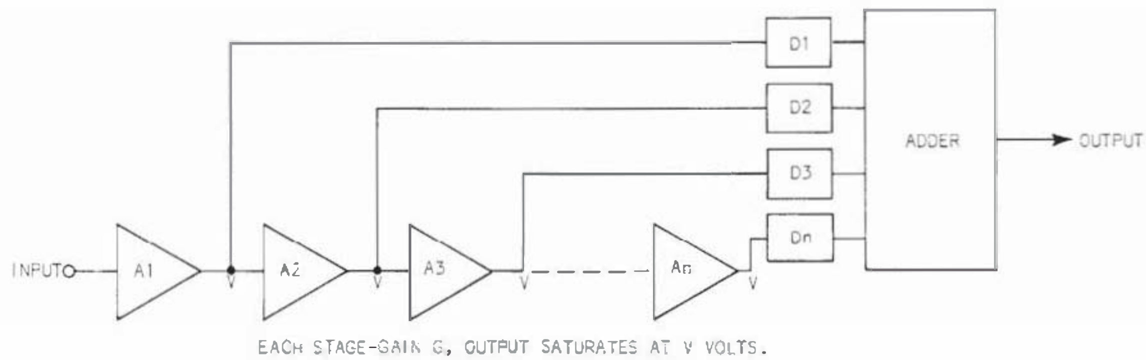


Fig. 4-9. Successive-detection logarithmic amplifier.

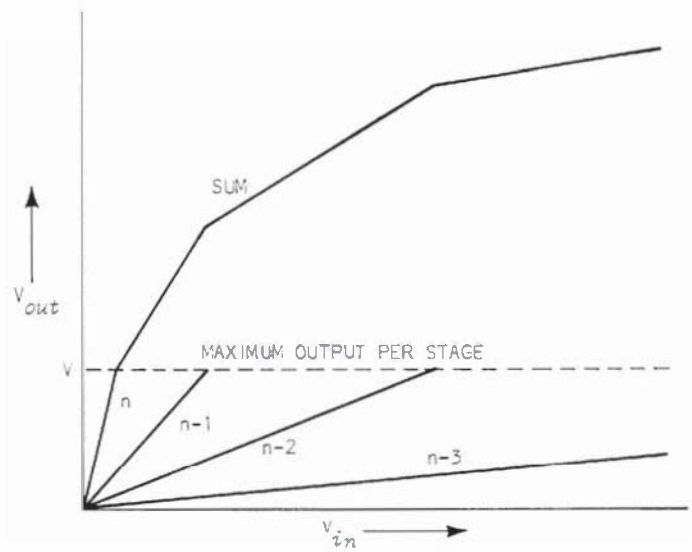


Fig. 4-10. Successive-detection logarithmic IF voltages.

successive detection

One of the early schemes for a linear-limiting log-amplifier is known as a *successive-detection* logarithmic amplifier. This system is based on a cascade of linear amplifier stages, each having its own detector as shown in Fig. 4-9.

last stage saturates first

The output of each stage goes to the next stage and to a separate linear detector. The outputs of the individual detectors are added and this constitutes the final output. At low signal levels the contributions of the early stages is negligible; the total output consists essentially of the contribution from the last stage. Finally we reach a signal level (V) at which the last stage (An) is saturated, its output does not increase with an increase in input. If the input is now increased by the factor G, the gain per stage, the (n-1) stage will saturate and produce an output V. The total output has now increased to 2V, while the input went up by a factor of G. This process can be continued for the other stages, and since the output increases in arithmetic progression while the input proceeds in geometric progression we have a logarithmic amplifier. The logarithmic properties of this amplifier are evident from an inspection of Fig. 4-10. The construction shows four stages each having a gain of four, so that the next to last (n-1) stage saturates at an input which is four times as large as that needed to saturate the last (n) stage. As more stages are added the break points smooth out, so for all intents and purposes we have a continuous logarithmic characteristic. However, for proper operation it is necessary that the detector characteristics, linear or square-law or in between, be precisely controlled -- a difficult task.

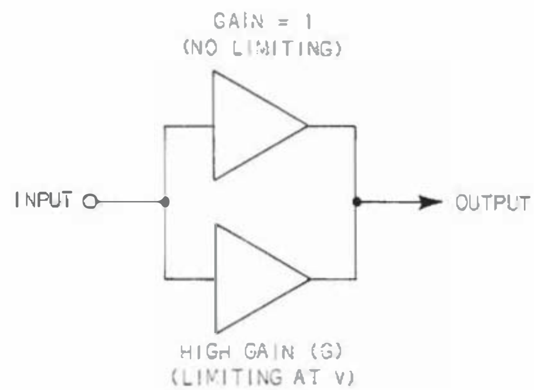


Fig. 4-11. Twin-gain amplifier stage.

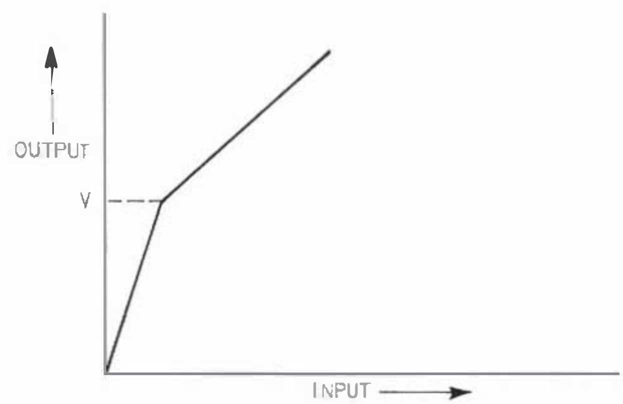


Fig. 4-12. Twin-gain amplifier-stage transfer characteristic.

twin-gain

In general the difficulties with the successive-detection method is that it only works in conjunction with a detector. Sometimes it is desirable to get a logarithmic response at IF frequency without detection of the signal. The *twin-gain* approach to this requirement was proposed in 1966. The amplifier is composed of stages having two modes of amplification. At low-level inputs we have a linear high-gain amplifier state, while at high-level inputs the gain changes to unity. Such an amplifier stage can be represented by the block diagram of Fig. 4-11. The input-output characteristic of such an amplifier stage is shown in Fig. 4-12. The output increases at the rate of "G" times the input till the critical voltage "V" is reached, at this point the output starts increasing at the same rate as the input. It will be noted that the major difference between a successive-detection stage and a twin-gain stage is that the former saturates (no increase in output with increasing input) upon reaching the critical output-voltage V, whereas the latter does not saturate but the gain drops to unity. A cascade of twin-gain stages will produce an input-output characteristic similar to that shown in Fig. 4-10, but with one major difference. In the twin-gain system all the action is at IF frequency and no detectors are involved.

REFERENCES; see page 171.

Amplifiers -- B-10, B-11, C-10
Low-Noise Amplifiers -- C-9
Logarithmic Amplifiers -- C-11

5

MIXERS

The superheterodyne technique on which Tektronix spectrum analyzers are based depends on a frequency-translation process. A mixer is a three-terminal device, having a signal (RF) input, a local-oscillator (LO) input, and an intermediate-frequency (IF) output. Ideally the mixer simply translates the input signal in frequency but leaves everything else (e.g., modulation, spectral distribution, etc.) unchanged. In practice the mixer introduces various aberrations such as intermodulation, spurious outputs, signal limiting, and others.

two inputs one output

In the sections that follow we shall discuss some of the characteristics of mixers in general, and then go on to more detailed descriptions of various specific types of mixers.

CHARACTERISTICS OF MIXERS

Fig. 5-1 shows a standard three-terminal mixer. We define the *conversion loss* (L) as the ratio of input level A to output level C, thus $L = \frac{A}{C}$. This definition assumes that input and output impedance levels are equal. Therefore, we can specify the conversion loss in decibels, thus $L_{dB} = 20 \log_{10} \frac{A}{C}$. For unequal impedance levels, conversion loss is specified in terms of power, thus

$$L_{dB} = 10 \log_{10} \frac{\text{power in (RF)}}{\text{power out (IF)}}.$$

conversion loss

dB = $10 \log \frac{W_{RF}}{W_{IF}}$

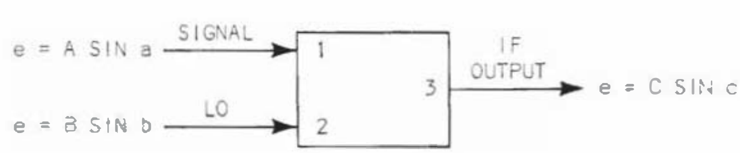


Fig. 5-1. Basic mixer configuration.

An ideal mixer has a conversion-loss ratio of unity or zero dB. Another parameter pertaining to signal levels is the *isolation* between terminals. Here the interest is to determine how much of the signal entering any terminal comes out of another terminal:

The LO-to-IF isolation is $I_{2,3} = 10 \log_{10} \frac{P_2}{P_3}$;

similarly $I_{2,1} = 10 \log_{10} \frac{P_2}{P_1}$ defines how much LO signal comes out of the signal-input terminal, and so forth. An ideal mixer would have infinite isolation. Other parameters pertaining to signal amplitude are associated with maximum-power-handling capabilities. This includes the *safe power level* pertaining to the maximum power that can be accommodated without physical damage (e.g., diode burnout), and *maximum input power* which refers to the maximum power that can be accommodated without degradation in performance. Like the IF amplifier, the mixer can introduce noise into the signal, thus degrading system sensitivity. The noise properties of a mixer are specified in terms of the *noise-temperature ratio*, (t). The noise-temperature ratio for a resistor is unity, so, except in some very special circumstances*, noise temperatures of mixers go from unity and up.

amplitude characteristics

Besides specifying the mixer parameters pertaining to signal level, it is also necessary to specify the parameters which pertain to frequency. Mixer-frequency specifications include: Operating frequency ranges (e.g., 1 MHz to 1 GHz) for each of the three terminals (RF, IF, LO); conversion-loss flatness as a function of frequency, and spurious-response characteristics. The spurious-response characteristics of a mixer are quite complex so that this specification calls for more than just a few numbers (see diode characteristics, Chapter 2). In general, a spurious-response specification would include image rejection, which is specified as

frequency characteristics

*The Schottky Barrier diode has a theoretical noise-temperature ratio which is less than unity. Whether or not mixers having less than unity noise-temperature ratios can be constructed is still debatable.

10 $\log_{10}$ of the power-output ratio of that at the image frequency to that at the desired frequency; local oscillator and signal feedthrough which is related to isolation; harmonic conversion efficiency, and the several other types of responses defined in the definitions section of Chapter 1.

In addition to the above it is frequently necessary to specify other operating conditions such as bias voltages and crystal current. In any event there are many different types of mixers, each type requiring somewhat different treatment as discussed in the following sections.

THE SINGLE-ENDED MIXER

a nonlinear element required

One of the simplest configurations is that of the single-ended mixer. Mixing is performed by a single nonlinear element such as a diode, as shown in Fig. 5-2. The function of the components is as follows: Capacitors C1, C2, and C3 are DC blocks to permit crystal-current control by means of R without reference to RF, LO, and IF source impedance.

optimize operating point

The diode is a nonlinear mixing element as discussed in Chapter 2. Radio-frequency-choke L1 and resistor R provide a direct-current path for the rectified local-oscillator signal which forms the crystal current. The crystal current sets the diode quiescent operating-point, which in turn determines some of the mixer operating parameters. The operating point can be optimized by adjustment of R or by application of external DC bias through R. Capacitor C2 is a small capacitor with negligible reactance at LO frequency but appreciable reactance at the much lower IF frequency. In terms of basic characteristics, C2 provides a certain amount of IF-to-LO isolation. Inductor L2 is a small choke which has negligible effect at IF frequency but prevents local-oscillator power from getting into the IF.

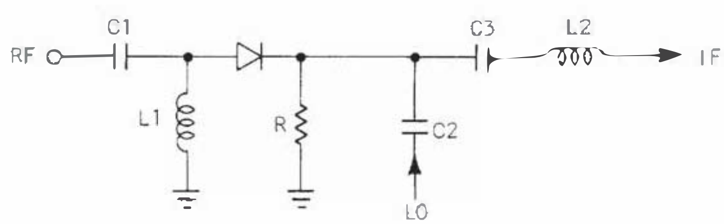


Fig. 5-2. Single-ended mixer.

Obviously the arrangement shown in Fig. 5-2 is one of many possible configurations. In particular it should be noted that the design shown in 5-2 is based on the assumption of a high local-oscillator frequency and a relatively low IF frequency (e.g., 1 GHz and 100 MHz). In many instances the local-oscillator frequency and the IF frequency are not much different; under these circumstances, we could not provide IF-LO isolation by using filtering techniques.

It should now be clear that single-ended mixers are used at many different frequencies, for a variety of applications, and are constructed in various ways. Fortunately there is no need to consider all of the variations separately. The single-ended mixer has certain basic performance characteristics regardless of peculiarities of design:

A) *Frequency Response* -- Single-ended mixers are essentially untuned broadband devices. A typical specification might be:
   RF-frequency range: 10 MHz to 12.4 GHz
   LO-frequency range: 100 MHz to 12.4 GHz
   IF-frequency range: 30 MHz to 100 MHz.
   Note that the IF-frequency range is considerably narrower than either the RF or LO range. This is necessary to provide some isolation between the circuits connected to the three terminals.

B) *Conversion Loss* -- In general the wider the operating-frequency range the greater is the conversion loss. This is so because broadband diodes have poorer conversion-loss properties than units optimized for a specific narrow frequency range. Likewise, broadband coupling structures usually exhibit more loss than an optimized narrowband arrangement. Frequently the broadband coupling structure will include a resistive attenuator to help improve the impedance match (VSWR) of the mixer. Naturally it is easier to obtain a good VSWR over a narrow frequency range than over a wide frequency range. Finally a broadband mixer having no spurious-response rejection will suffer in conversion efficiency due to the energy that is lost in the image and other spurious responses. Typical

conversion-loss numbers would be about 3 dB for a narrowband highly optimized unit to 15 dB* for a broadband mixer.

C) *Noise* -- The noise contributed by the mixer is fairly low, and is usually of little consequence in sensitivity computations. Typical noise-temperature ratios are about 1.5. Of greater consequence is noise from the local oscillator getting in via the mixer. This effect can be reduced by improved local-oscillator design and the use of a balanced mixer.

D) *Impedance Match* -- Single-ended broadband mixers usually exhibit poor input-impedance properties; a VSWR of 4:1 is typical. The input impedance of a mixer depends on many parameters: The properties of the diode, bias setting and local-oscillator power level. As a result, it is almost impossible to obtain a good VSWR by lossless-matching techniques. The standard method is to add a 3- to 6-dB attenuator for VSWR improvement when a poor VSWR cannot be tolerated. Naturally this degrades sensitivity by the amount of added signal attenuation.

E) *Spurious Responses* -- The broadband single-ended mixer configuration does not provide spurious-response rejection. As a result, the image and all the harmonic conversions appear at the IF output terminal. The frequency relationship between the RF input signal at f_{RF}, the local-oscillator input at f_{LO}, and the IF output at f_{IF} is given by:

image and harmonics

$$\left.\begin{aligned} M\, f_{RF} \pm N\, f_{LO} &= f_{IF} \\ N\, f_{LO} \pm M\, f_{RF} &= f_{IF} \end{aligned}\right\} \text{ M \& N positive integers.}$$

---

* This does not include the additional losses due to operation on local-oscillator harmonics. See section on harmonic mixers for more details.

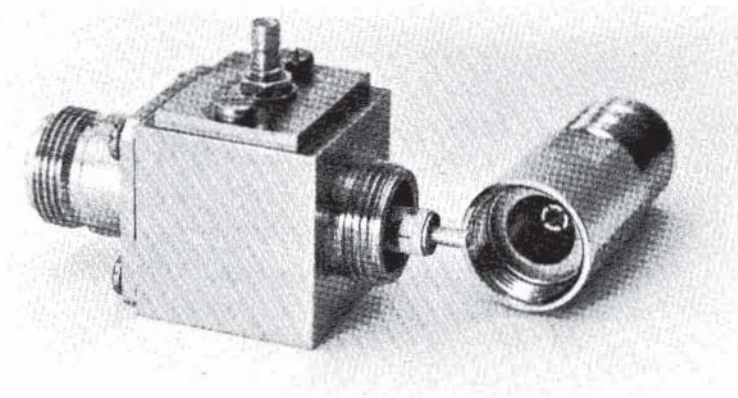


Fig. 5-3. Single-ended mixer.

relative amplitude of spurii

The relative amplitude of the various responses can be approximately computed as indicated in Chapter 2 on diodes. In general the effect of signal harmonics ($M \geq 2$) is negligible unless the signal is quite large. The same holds true for intermodulation. Permissible input power levels are defined by the *Maximum Input Power* specification. Effects due to local-oscillator harmonics can be controlled, to some extent, by judicious choise of bias. However, very little can be done about the image (i.e., $f_{RF} - f_{LO}$, or $f_{LO} - f_{RF} = f_{IF}$) or IF feedthrough (i.e., $f_{LO} = f_{IF}$, $f_{RF} = f_{IF}$).

configuration options

Single-ended mixer construction varies, depending mainly on the frequencies involved. At relatively low frequencies the mixer can be constructed in almost any configuration. At high frequencies (hundreds of MHz), however, the mixer is usually constructed in a transmission-line structure. Fig. 5-3 is a photograph of a coaxial single-ended mixer.

BALANCED MIXERS

differing degrees of balance

Balanced mixers provide isolation between the several terminals. Mixers have differing degrees of balance depending on whether only two, or more than two, terminals are isolated from each other. Note that there are six possible combinations, namely: RF - IF, RF - LO, IF - LO, and two combinations of two at a time and one combination of all terminals isolated from all other terminals.

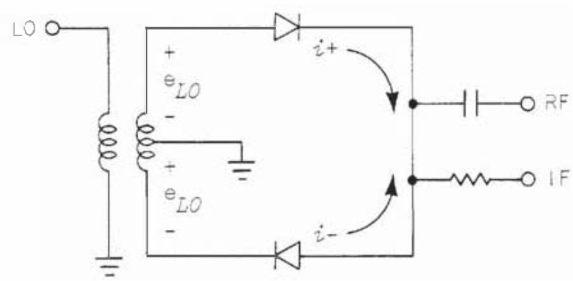


Fig. 5-4. Schematic of a single-balanced mixer.

It is necessary to have a minimum of one balanced circuit to achieve minimal isolation: One terminal from one other terminal. This is called a single-balanced mixer. To achieve isolation between all the terminals it is necessary to add only one other circuit. Thus, our interest stops at double-balanced mixers.

Balanced mixers can be constructed in many different configurations. Fig. 5-4 shows one type of single-balanced mixer.

transformer balance

Assuming that the diodes are identical and the input transformer perfectly center-tapped, no local oscillator energy should get into the IF or RF terminals. This is evident from the observation that the input transformer splits the local-oscillator input into two identical circuits whose outputs combine at the IF output terminal. Since $i+$ and $i-$ are of equal magnitude but opposite sense they cancel each other. Mixing occurs when the diode balance is disturbed by the application of RF.

LO-to-IF balance

balancing controls

Theoretically, as described above, LO to IF-RF balance should be perfect (infinite isolation). Actually the balance can be quite poor (e.g., 10-dB isolation). This is especially true if the mixer operates over more than a very narrow frequency range. The reason is that it is difficult (or very expensive, which is equivalent to difficult) to get well-matched diodes, and perfectly tapped transformers. Further the mixer is built on a chassis, and that means stray capacitance to ground. As a result practical mixers include balancing controls, such as variable resistors or capacitors. These controls are used to compensate for circuit unbalance by adjusting the relative LO level getting into the mixer diodes.

two-diode and four-diode mixers

The mixer configuration gets more complicated when one gets to double-balanced mixers. Though there are many variations, those can all be classified into one of two basic systems. These are the *two-diode* or *series-balanced* system and the *four-diode* or *ring-modulator* system*. Fig. 5-5 and 5-6 illustrate these basic systems. The balancing effect is best illustrated with reference to the equivalent circuits of Figs. 5-7 and 5-8. In the two-diode modulator we have LO-RF and LO-IF isolation, while the RF signal will appear at the IF-output terminal. In the equivalent circuit of Fig. 5-7 we have the RF signal split into two equal components (e_s). With the mixer-diode polarity as shown, the RF input appears full-wave rectified at the IF output, and there is no IF-RF isolation.

Contrast this with the case in Fig. 5-4, where the LO is injected in the same manner as the signal in Fig. 5-5 and isolation is obtained by inverting the polarity of one of the diodes. With local-oscillator polarity as shown, i_1 and i_2, equal amplitude and opposite sense, flow into a common junction and cancel. No local-oscillator current flows during the other half cycle. When both local oscillator and RF signal are present at the same time, the diode balance is disturbed and currents at combination frequencies appear at the output.

Let us now consider the most complicated case, namely the four-diode mixer. Here, unlike the two-diode case, current flows in the primary of T2 during both half cycles of local-oscillator input.

signal and harmonics disappear

During one half cycle, current flows toward the center tap (i_1 and i_2), and during the second half cycle, current flows away from the center tap (i_3 and i_4). Cancellation occurs during each half cycle in the same manner as for the two-diode mixer. One advantage of the four-diode mixer is that the signal and all its harmonics disappear from the IF output (e_{IF}). This can be surmised from the equivalent circuit (Fig. 5-8), where it will be observed that if the diodes are assumed to have negligible forward resistance, all of the current due to e_s will flow through the diodes and none through the load resistors, R.

*Some classifications consider the two-diode mixer singly balanced since, like the mixer of Fig. 5-4, there is no RF-to-IF isolation. The choice of classification in this volume is arbitrary.

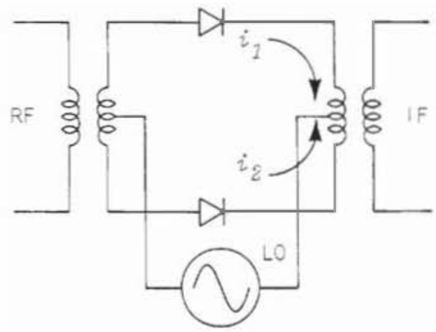


Fig. 5-5. Double-balanced two-diode mixer.

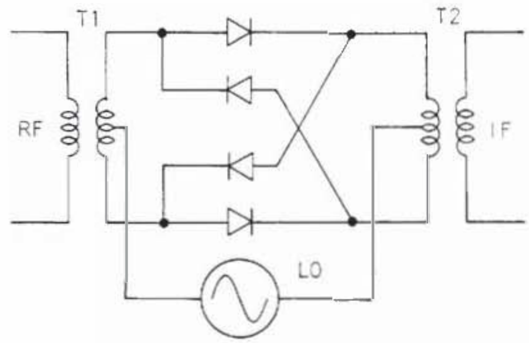


Fig. 5-6. Double-balanced four-diode (ring-modulator) mixer.

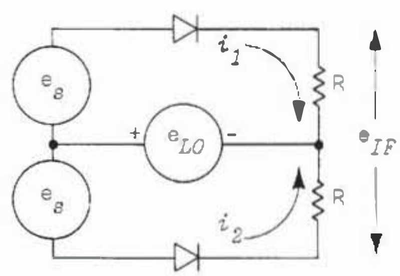


Fig. 5-7. Two-diode balanced-mixer equivalent circuit.

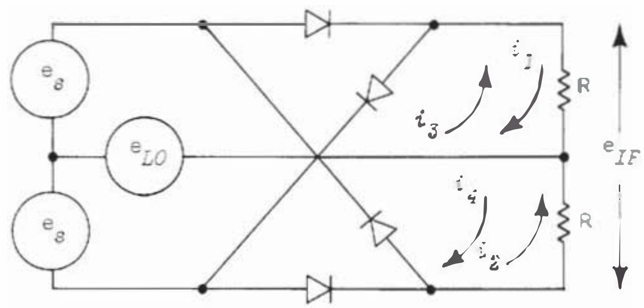


Fig. 5-8. Four-diode balanced-mixer equivalent circuit.

Besides the physical arguments presented above, one can obtain a complete mathematical solution by combining the standard mesh equations of Kirchhoff's laws with the series representation of the diode V-i characteristic given by Equation 2 in diode characteristics section of Chapter 2.

Table 5-1 gives a comparison of the IF output frequencies, up to the fifth harmonic, for the two-diode and four-diode systems.

MIXING PRODUCT	TWO DIODE	FOUR DIODE
f_{RF}	•	
$2f_{RF}$		
$3f_{RF}$	•	
$4f_{RF}$		
$5f_{RF}$	•	
f_{LO}		
$f_{LO} \pm f_{RF}$	•	•
$f_{LO} \pm 2f_{RF}$		
$f_{LO} \pm 3f_{RF}$	•	•
$f_{LO} \pm 4f_{RF}$		
$f_{LO} \pm 5f_{RF}$	•	•
$2f_{LO}$		
$2f_{LO} \pm f_{RF}$	•	
$2f_{LO} \pm 2f_{RF}$		
$2f_{LO} \pm 3f_{RF}$	•	
$2f_{LO} \pm 4f_{RF}$		
$3f_{LO}$		
$3f_{LO} \pm f_{RF}$	•	•
$3f_{LO} \pm 2f_{RF}$		
$3f_{LO} \pm 3f_{RF}$	•	•
$4f_{LO}$		
$4f_{LO} \pm f_{RF}$	•	
$4f_{LO} \pm 2f_{RF}$		
$5f_{LO}$		
$5f_{LO} \pm f_{RF}$	•	•

f_{RF} = RF INPUT SIGNAL AT FREQUENCY f_{RF}
f_{LO} = LOCAL-OSCILLATOR DRIVE AT f_{LO}
• = OUTPUT CONTAINS THIS PRODUCT

Table 5-1. Mixer-output products.

The advantages of the four-diode system over the two-diode system in suppressing undesired modulation products is obvious from an examination of Table 5-1. It should be noted that both the two-diode and four-diode mixer to not suppress the odd-order products (e.g., $f_{LO} \pm 3f_{RF}$). This means that both types of mixer, if driven sufficiently hard by two RF signals, will produce odd-order intermodulation products.

intermodulation products

As with the single-balanced mixer, an actual double-balanced mixer would contain balance adjustments since it is impossible (uneconomical) to perfectly match the mixer components. Furthermore, an actual mixer might contain matching resistors, or transmission-line transformers for broadband coupling, or two sets of balance controls -- one for amplitude and one for phase; or some other variations. The important thing, in trying to understand how it works, is to separate the essential components common to all mixers from the specialized additions that give a mixer its individual character. This is illustrated by Fig. 5-9, where T is a transmission-line transformer, Rl is a matching resistor, R2 is an amplitude balance, and Cl and C2 affect both amplitude and phase. In spite of these variations this is a single-balanced mixer as is that appearing in Fig. 5-4. Fig. 5-10 is a photograph of such a mixer.

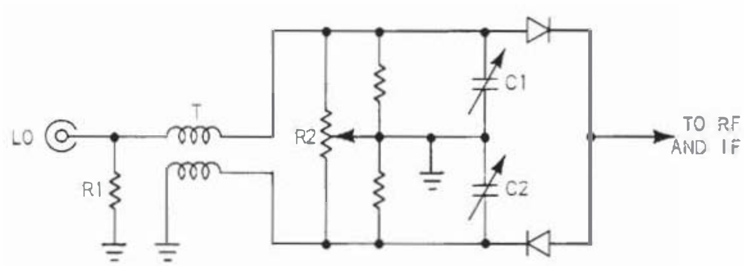


Fig. 5-9. Single-balanced mixer with balance controls.

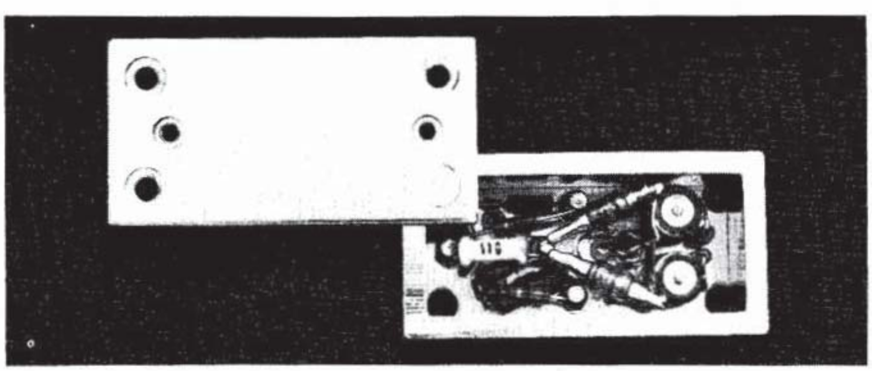


Fig. 5-10. Single-balanced mixer.

The performance parameters of balanced mixers vary depending on mixer design:

general characteristics

A) *Frequency Response* -- Balanced mixers are inherently narrower band devices than single-ended units. The limiting factor is usually the balanced-coupling structure. Typical RF-input frequency ranges would be 1-to-1000 MHz, 2-to-5 GHz, 12.4-to-14.4 GHz, etc.

B) *Conversion Loss* -- As with single-ended mixers, best conversion efficiency is obtained when the mixer is optimized for a narrow frequency range. Typical conversion-loss numbers are 3-to-10 dB, with most commercially available mixers in the 5-to-8-dB area.

C) *Noise* -- The noise contribution of the diodes has always been quite low, and with the recent introduction of the Schottky barrier diode, diode noise has become negligible.

LO noise

Besides diode noise one must also consider local-oscillator noise. Here the balanced mixer has an advantage over the single-ended unit. Consider, for example, a typical case: LO frequency = 1 GHz, IF frequency = 100 MHz, LO power = 10 mW: The noise power from a noisy oscillator, integrated over the IF bandwidth is 100 dB below the carrier power at 100 MHz from the carrier. This means that besides feeding a 10-mW 1-GHz signal into the LO terminal we are, among other things, also feeding two -90 dBm signals at 0.9 GHz and 1.1 GHz into the LO terminal. In an unbalanced mixer these noise signals combine with the 1-GHz LO to produce a 100-MHz IF output. Obviously the sensitivity of the system has been degraded. In a mixer having LO-IF isolation this kind of conversion cannot take place, and the noise performance of the system is improved.

better impedance match

D) *Impedance Match* -- Since balanced mixers usually operate over narrower frequency ranges than single-ended units, they tend to have a better impedance match. There are, however, applications where a good VSWR is of such importance as to require the addition of resistive attenuators for matching purposes.

E) *Spurious Responses* -- Balanced mixers generate fewer undesired outputs than the single-ended configuration, as illustrated in Table 5-1. The major improvement is in the suppression of IF-feedthrough signal. However, the image remains. Techniques for constructing imageless-mixers will be discussed in a later section.

HARMONIC MIXERS

A mixer which is optimized to produce an IF output by combining a harmonic of the local-oscillator drive with the RF input is called a *harmonic mixer*. In mathematical terminology we are working with:

$$\left.\begin{aligned} f_{RF} \pm N\, f_{LO} &= f_{IF} \\ N\, f_{LO} \pm f_{RF} &= f_{IF} \end{aligned}\right\} \quad \text{Where N is an integer greater than 1.}$$

Essentially we are utilizing some of our spurious responses for a useful purpose. Of course, we no longer consider these outputs spurious. Harmonic mixers are frequently used for frequency-range extension. Operation up to the tenth LO harmonic is not uncommon.

single-ended for harmonics

It is evident from Table 5-1 that most balanced configurations make for poor harmonic mixers. We shall, therefore, concentrate on a discussion of single-ended mixers when operated at local-oscillator harmonics.

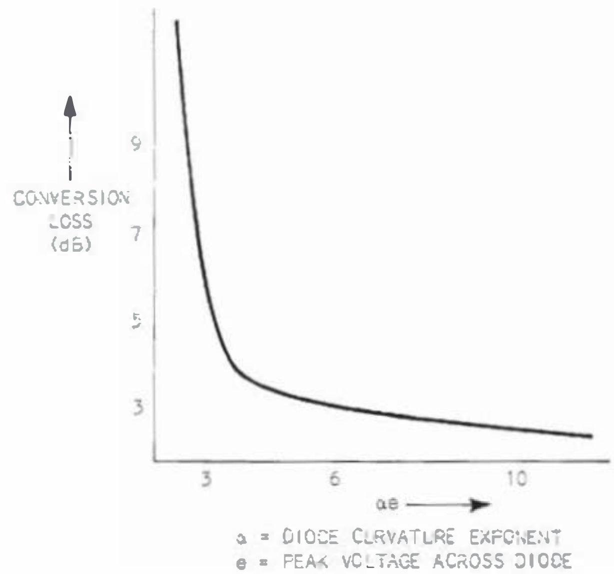


Fig. 5-11. Conversion loss vs local-oscillator drive.

The conversion loss of a single-diode mixer operating at LO fundamental decreases asymptotically with increasing local-oscillator drive. From Fig. 5-11 it appears that αe should be greater than about 3. The upper limit is established around $\alpha e \approx 10$, since higher drive levels increase diode noise and reduce diode reliability. The significance of the diode characteristic, α, is discussed in Chapter 2.

reduce crystal current

Harmonic mixers are usually operated on the same LO-drive levels as fundamental mixers. The basic difference between the two is in the rectified crystal-current setting. Fundamental mixers work best at full crystal current, that is with no diode back bias. Harmonic mixers, on the other hand, have best efficiency when crystal current is reduced by back bias. The reason for this can be surmised from the basic diode equations presented in Chapter 2.

high α^N for high LO harmonic

The higher the LO harmonic at which mixing is desired the higher the degree of the nonlinearity coefficient (α^N where N = harmonic number) that has to be optimized, and each of these gets optimized at a different quiescent operating point.

harmonics versus crystal current

Another way of looking at it is this. Not only do we need an efficient mixer, but we also need an efficient harmonic generator to generate LO harmonics which will then be mixed with the incoming RF signal. Unlike the basic mixing process where conversion efficiency is reasonably independent of LO power over broad power ranges, harmonic-generation efficiency is highly dependent on crystal current which is in turn dependent on LO power. The result is that it is difficult to maintain optimum harmonic-conversion efficiency with LO-power variations without readjusting the crystal current. There are two ways out. One is to maintain constant LO power, the other is to provide a means for reoptimizing the mixer as needed. The mixer optimizing technique is called *mixer peaking*. It consists of a means for controlling the rectified crystal current. This can be a resistor such as R in Fig. 5-2 or some other method.

mixer peaking

Harmonic mixing efficiency is always poorer than that for fundamental mixing. Average increase in loss is about 3 dB per harmonic with mixer peaking optimized. Thus, a mixer operating on the 10th harmonic of the local oscillator would have about 30 dB more conversion loss than the same mixer working on fundamental.

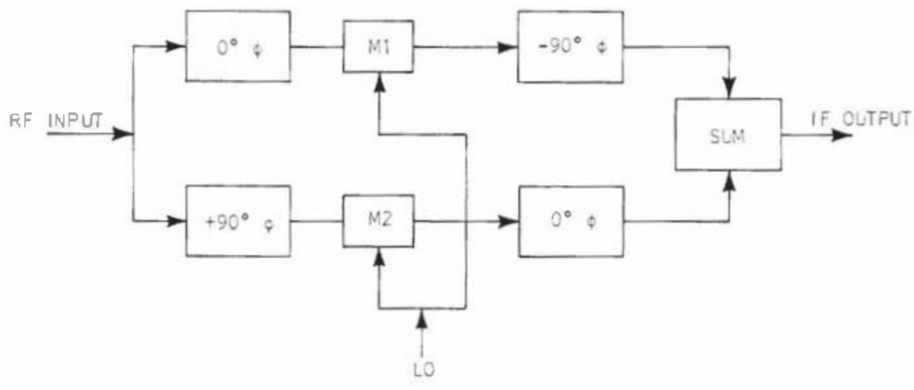


Fig. 5-12. Imageless-mixer block diagram.

MISCELLANEOUS MIXERS

Among the various categories of mixers that might be encountered in spectrum analyzers are: Imageless mixers, waveguide mixers, high-output-frequency mixers, switching mixers, mixers that use diplexers, and nondiode mixers. We shall consider these in turn.

remove image

sum and difference

A) *Imageless Mixers* -- One of the major spurious responses in a superheterodyne receiver or spectrum analyzer is the image. The image is very troublesome since the mixer-output level is attenuated very little, if at all, with respect to the desired signal. The image can, however, be eliminated by the two-mixer and phase-shifter arrangement shown in Fig. 5-12. The system is based on the frequency relationship between the desired IF output and the image: Assume that the desired output is f_{IF} = E cos $(\omega_{LO} - \omega_{RF})$; then the image f_{IF} = E cos $(\omega_{RF} - \omega_{LO})$. In each conversion the frequency and relative phase information of the signal (ω_{RF}) is preserved. Thus, for mixer number 2 relative to mixer number 1, the desired signal is shifted by -90° while the image is shifted by +90°. Mixer 1 output is now shifted by -90° with respect to mixer 2. The input to

the summation network consists of both signal and image at -90° from mixer 1, and signal at -90° and image at +90° from mixer 2. The summing network adds the two signals which are in-phase (both at -90°), and cancels the image components which are 180° out of phase.

There are several variations on this basic system; for example, one uses phase-shift networks in the local-oscillator line. The basic performance parameters are, however, the same. The reason why such mixers are seldom used is the difficulty of obtaining good balance. The image cancellation is highly sensitive to both amplitude and phase differences between the two channels. As a result such mixers are used over relatively narrow frequency ranges, and where an image cancellation of about 40 dB is adequate.

B) *Waveguide Mixers* -- In a waveguide mixer the signal and/or the local-oscillator energy is propagated in waveguide. The basic mixer theory and varieties of mixer configuration all apply to waveguide mixers. The major differences between waveguide mixers and other mixers are in construction technique -- dictated by the use of waveguide, and the peculiarities due to a high operating frequency since waveguide is seldom used at low frequencies (waveguide is usually used above 12.4 GHz).

higher conversion loss at higher frequency

High-frequency operation requires an increased use of harmonic mixers with an attendant decrease in sensitivity. Furthermore, the crystal diodes available at these frequencies are less efficient than their low-frequency counterparts. The result is that waveguide mixers usually have higher conversion loss than coaxial units. *It should be kept in mind that this difference is not due to the use of waveguide or coax, but rather to the difference in operating frequency.*

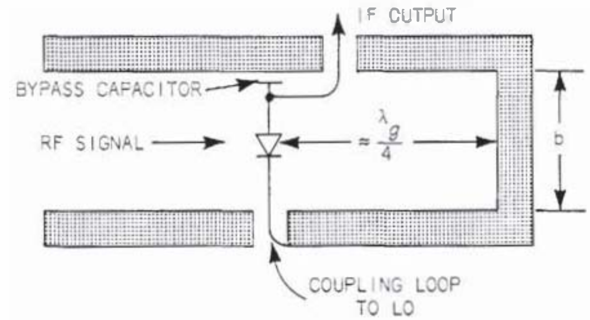


Fig. 5-13. Waveguide-mixer construction.

As with other types, there are many ways of constructing waveguide mixers. Fig. 5-13 shows a basic waveguide-mixer construction technique. The diode is connected across the narrow dimension of the waveguide (b). Local-oscillator signal is injected through a coupling loop while the IF output is filtered out by adding a small capacitor to ground. The RF signal propagates down the waveguide, and is prevented from radiating out by having the mixer closed at the opposite end. The shorting wall is positioned approximately a quarter guide wavelength from the diode so that it reflects a high impedance at the plane where the diode is situated. Obviously this is not a broadband mixer, since the short will reflect an increasingly lower impedance as the distance between short and diode departs from $\frac{\lambda_g}{4}$. Other variations use tapers, steps, or other means for improving the broadband properties of the mixer. Fig. 5-14 is a photograph of a waveguide mixer. Here the diode sits outside the waveguide in a coaxial structure. The center conductor of this coax continues on across the waveguide and connects to the output filter. Local-oscillator power is injected through the IF-output terminal by means of a diplexer.

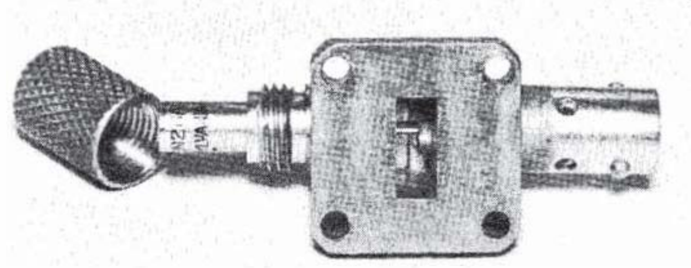


Fig. 5-14. Waveguide mixer. Courtesy of Sage Laboratories.

different approach for high-IF-frequency mixers

C) *High-Output-Frequency Mixers* -- When one speaks of mixers there is the tacit assumption that the IF-output frequency is lower than the RF and/or LO frequency. This is not always the case. There are situations where, to improve image separation or for other reasons, a high IF frequency is used. This creates signal separation problems, since the standard technique of using low-pass filters in the IF line can no longer be used. As a result most high-IF-frequency mixers (IF frequency in the GHz region) use one or more of the following: directional couplers, directional filter, diplexers, high-Q bandpass or bandstop filters.

conserving IF power

A typical configuration illustrating the problems of high-IF-frequency mixers is shown in Fig. 5-15. The directional coupler couples at local-oscillator frequency (e.g., 2-to-4 GHz) but not at IF frequency (e.g., 1 GHz). The output tank circuit is resonant at IF frequency, so most of the converted energy at IF frequency is coupled through this low-loss circuit to the IF output. The band-reject filter at the RF input serves two purposes. First, it reduces the IF-feedthrough spurious response. Second, any IF frequency signal that travels down the transmission line toward the RF input gets reflected back toward the IF output so that conversion efficiency is improved. When the local-oscillator frequency range is within the RF-input frequency range it is necessary to prevent loss in sensitivity by the RF signal coupling into the directional-coupler termination.

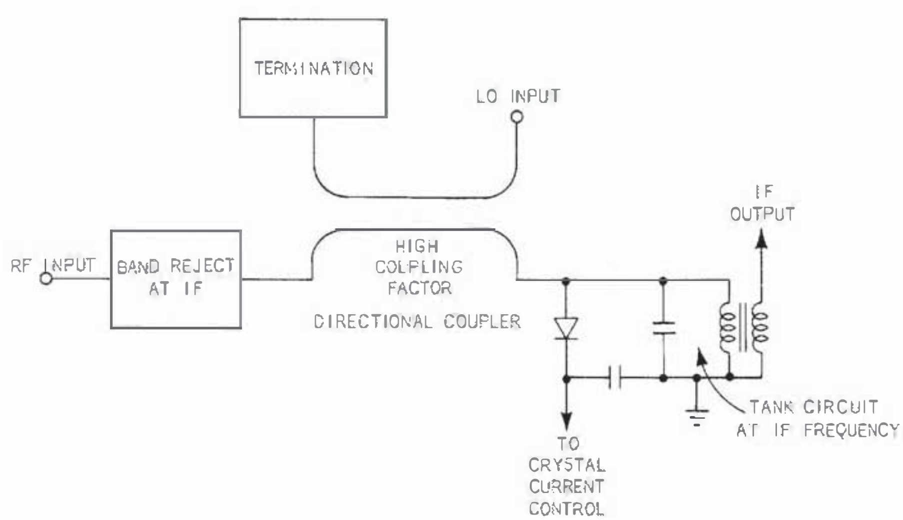


Fig. 5-15. High-IF-frequency mixer.

This is usually accomplished by using a high-coupling-factor (e.g., 10 dB) directional coupler. This necessitates the use of a higher-power local oscillator since a smaller portion of LO power gets to the diode.

D) *Switching Mixers* -- As indicated in Chapter 2, it is theoretically possible to construct a lossless mixer by the use of perfect switches. Of course, there is no such thing as a perfect switch. No diode has infinite back resistance or zero forward resistance. Actually the back resistance of most mixer diodes is sufficiently high to be of minor concern; it is the forward resistance that causes most of the losses. Recently, with the introduction of the low R_s (about 10 Ω) Schottky barrier diode, it has become possible to construct true switching-mode mixers.

low forward R for switching mixers

Obviously, if the mixer diode behaves as a perfect switch, there can be no losses in the diode. Thus, mixer losses can come about from the following sources only: 1) Loss of RF signal in the RF-to-diode coupling structure or RF signal getting into the IF and LO terminals, 2) Loss of IF signal in the IF-out-to-diode terminal or IF signal getting into the RF and LO terminals, 3) Loss of converted signal in the image or other spurious responses.

To get a good approximation to a switch it is necessary to use a low-forward-resistance diode that is switched by high local-oscillator power (e.g., 50 mW). Thus, the requirements for a good switch have no effect on the basic structure of the mixer. The circuit losses, however, can only be reduced by judicious design procedure as demonstrated in Fig. 5-16. The first thing to recognize is that this is a narrowband mixer. The RF, LO, IF, and image frequencies are all fixed. This permits the use of high-Q low-loss circuits. The high-Q RF and IF filters prevent converted signals from going anywhere except to the IF output. The image-frequency trap is adjusted to reflect the image-frequency energy back into the diode at an appropriate phase angle to reconvert (or remix) it to IF frequency. The phase properties of the image-frequency trap are very critical, so the circuit usually has

fixed-frequency operation

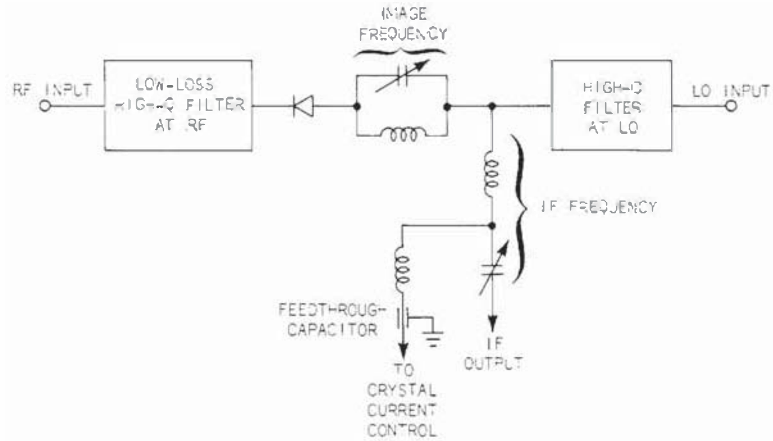


Fig. 5-16. Switching mixer.

to be adjustable. The desired signal is filtered by the series LC circuit and delivered to the IF amplifier. This type of mixer when properly constructed can have a conversion loss of less than 2 dB into an IF in the GHz region.

separate the diode from circuit environment

E) *Mixers Containing Diplexers* -- As previously indicated, the diplexer can be used for LO-IF separation in a mixer. The main advantage of this system is that the mixing element (e.g., crystal diode) can be separated from the main body of the instrument. Interconnection requires a single transmission line. This permits easy interchange of mixing elements, and makes for a convenient physical configuration.

The crystal-diode mixing element is housed in a holder, usually coaxial or waveguide. The RF-input side contains a broadband crystal-current return such as a choke or resistive attenuator. The other side of the diode holder is a relatively broadband structure that will pass both the IF frequency and LO frequencies. Signal-filtering, and mixer-peaking control is taken care of by the diplexer. The complete mixer is shown in Fig. 5-17.

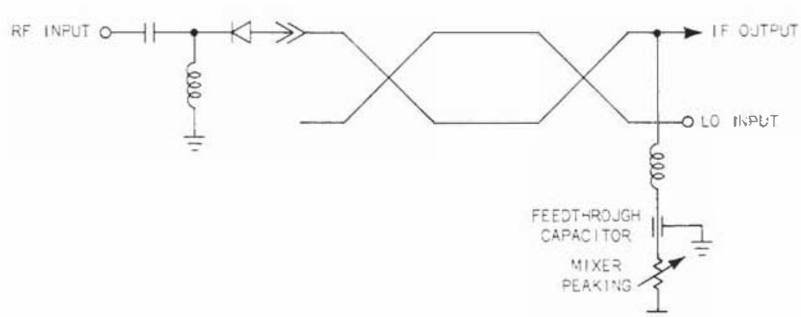


Fig. 5-17. Mixer using diplexer.

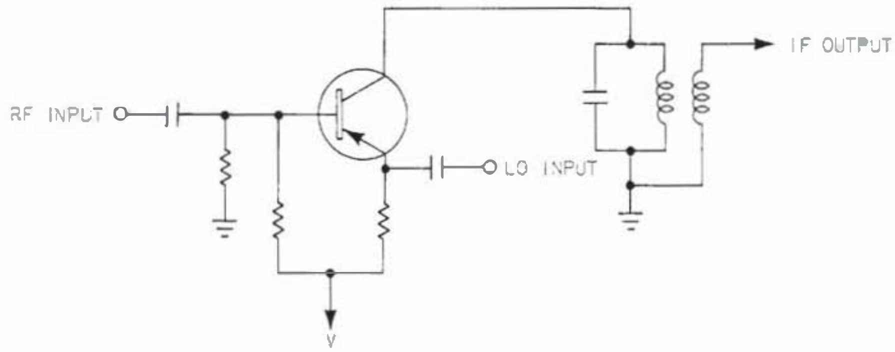


Fig. 5-18. Transistor mixer.

active-element mixer provides gain

F) *Nondiode Mixers* -- All of the mixers discussed so far use a crystal diode as the mixing element, because the crystal-diode mixer is used in a great majority of spectrum-analyzer applications. It should, however, be recognized that other types of mixing elements can be used. This would include both vacuum tubes and transistors. The advantage of using an active-mixer element is that we can also get gain along with the mixing. Thus, in a transistor mixer we get a conversion gain as opposed to a conversion loss. Fig. 5-18 shows a typical transistor mixer. The output tank circuit is tuned to IF frequency, thus eliminating RF and LO from the output. One can think of this type of mixer as a diode mixer in the base-emitter junction combined with a transistor amplifier.

REFERENCES; see page 171.

Mixers -- B-8, B-9, B-11
Balanced Mixers -- C-7 Chapter 5, pp98-106
Harmonic Mixers -- C-12
Imageless Mixers -- C-13

6

OSCILLATORS

oscillator categories

All frequency-translating spectrum analyzers contain at least one oscillator; some contain five or more. Oscillators fall in many categories -- tunable or fixed, microwave or low frequency, crystal controlled or phase locked or free running, electronically swept or mechanically tuned, electron tube or transistor or BWO or avalanche diode, and many other categories. All oscillators, regardless of type, have certain specifiable parameters, such as frequency range, power output, stability, etc. in common. We shall first consider these common specifiable parameters, and then go on to a discussion of specific types of oscillators.

BASIC OSCILLATOR PARAMETERS

oscillator operating-frequency range

The first point of interest, when discussing an oscillator, is usually the frequency of operation. This takes the form of a number such as 1 MHz or 3.5 GHz. Sometimes the oscillator is tunable; the operating frequency can be changed. This change can be performed by either electrical and/or mechanical means. When an oscillator is tuned we speak of an operating-frequency range which is characterized by two numbers, such as 3.5 GHz to 4 GHz. When we speak of a 3.5-GHz oscillator we do not mean that the frequency is precisely 3.5 GHz. What is meant is that the frequency is 3.5 GHz within certain limits. The accuracy of the frequency specification is usually given as a percentage of basic frequency, such as 3.5 GHz ± 1% at 25° C ± 100 parts-per-million/°C.

After establishing the basic frequency parameters, the next point of interest is usually power output. At relatively low frequencies (kHz and low MHz) it is not unusual to speak of voltage rather than power output. However, at microwave frequencies the discussion is almost invariably in terms of power, primarily because it is difficult to measure voltage and fairly easy to measure power. Power is specified in watts or milliwatts and in dBm (dB relative to one milliwatt). It is always understood that the power is delivered to a matched load, and that unless otherwise specified the load is 50 Ω. A power-output specification might be 20 mW (or +13 dBm).

output measured in power or dBm

50-Ω load

Once the basic frequency and output power specifications are established it becomes necessary to inquire into the effects of various operating conditions such as vibration, temperature, power-supply voltages, etc. This leads to specifications on incidental FM backlash, distortion, etc.

environmental influences

The following is a list of oscillator terms and specifications:

backlash -- The oscillator frequency remains unchanged for some tuning interval before reversing, when the oscillator tuning is reversed.

distortion, sinewave -- See *sinewave distortion.*

drift (frequency drift, stability) -- See *stability.*

FM'ing -- See *incidental frequency modulation.*

frequency range -- That range of frequencies over which a device meets its specification.

frequency reversal -- The oscillator-frequency change reverses direction, when the oscillator tuning is continuous in one direction.

frequency skipping -- When the oscillator tuning is smooth and continuous and the oscillator-frequency change is in discrete steps.

front lash -- The oscillator-frequency change continues in the original direction for some tuning interval before reversing, when the oscillator tuning is reversed.

incidental frequency modulation -- Short-term frequency jitter or undesired frequency deviation caused by instabilities in the spectrum-analyzer local oscillators.

micro -- See *microphonics*.

microphonics -- Incidental frequency modulation caused by mechanical vibration or shock.

mode shift -- Change in an oscillator's mode of operation usually indicated by a frequency skip or discrete amplitude change.

over-range -- That range of operation, beyond the frequency range, for which certain performance may be described.

sinewave distortion -- The measure of harmonic content of the oscillator output waveform into a specified termination.

spectral purity -- The measure of nonharmonic content of the oscillator output waveform into a specific termination.

squegging oscillator -- An oscillator that is self-pulsed.

stability -- Property of retaining defined electrical characteristics for a prescribed period. Deviations from a stable state may be called *drift* or *jitter*. In triggered-sweep systems, *triggering stability* may refer to the ability of the trigger and sweep systems to maintain jitter-free display of high-frequency waveforms for long periods of time (seconds to hours). Also, the name of the control used on some instruments to adjust the sweep for triggered, free-running, or synchronized operation.

CRYSTAL OSCILLATORS

stable oscillators required

Crystal oscillators are used in spectrum-analyzer circuits as reference sources in phase-lock systems, and as fixed-frequency local oscillators in the superheterodyne frequency-conversion chain. The overall stability of a spectrum analyzer is determined by the combined stability of the local oscillators. This leads to the use of stable high-Q resonators in the form of quartz crystals.

crystal replaces LC resonant circuit

Basically the crystal oscillator can be considered as a modified LC oscillator where one of the resonant circuits is replaced by the crystal or a resonant circuit has been added in the form of a crystal. Thus, most crystal oscillators are classified as a variation of some standard LC circuit such as Colpitts or Hartley.

$\omega^2 = \frac{C1 + C2}{L \cdot C1 \cdot C2}$

Crystal oscillators can be constructed in both vacuum tube and transistor configurations. However, Tektronix spectrum analyzers use transistor circuits only. The most popular circuit is a variation of the Colpitts oscillator as shown in Figs. 6-1 and 6-2. The circuit values for oscillation are computed from $\omega^2 = \frac{C1 + C2}{L \cdot C1 \cdot C2}$, in other words the inductor L is resonant with the capacitors C1 and C2 considered in series.

stability rule of thumb: C2 = 3C1

The ratio of C2 to C1 determines the stability properties of the oscillator. Designers have found that a ratio of three is usually optimum. Typical element values for a 1-MHz oscillator are L ≈ 100 μH, C1 ≈ 270 pF, C2 ≈ 1000 pF. As a matter of practical operation, one of the circuit elements, usually L, is made variable. This permits circuit optimization at crystal frequency. The circuit will not oscillate unless the relationship $\omega^2 \approx \frac{C1 + C2}{L \cdot C1 \cdot C2}$ is satisfied at the crystal frequency. Figs. 6-2A and 6-2B become identical at the frequency where the crystal is series resonant. Actually the above relationship need not be met precisely. The circuit is also tolerant of relatively high (several hundred ohms) series crystal resistance.

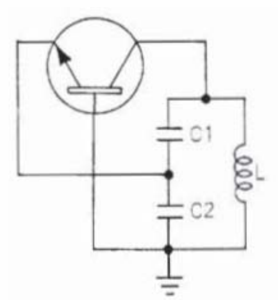


Fig. 6-1. Basic Colpitts oscillator.

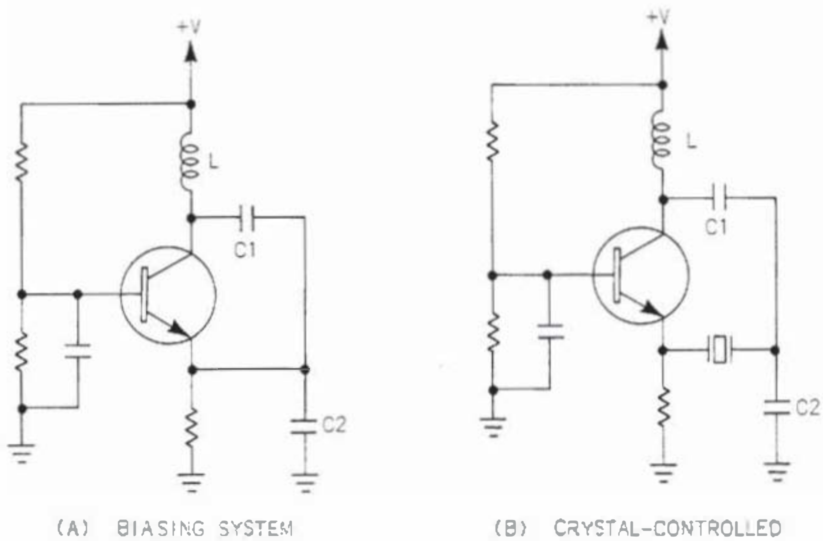


Fig. 6-2. Colpitts oscillator.

neutralize C_0 with a shunt inductor

Crystal oscillators of this type, using overtone crystals, can be used successfully in the 100-MHz region, provided that the crystal-shunt-capacity C_0 is neutralized. The effect of C_0 must be neutralized, since at 100 MHz a 10-pF C_0 appears as an impedance of less than 500 Ω across the crystal. The neutralizing element is usually a shunt inductor of appropriate value to form a high-impedance tank circuit with C_0.

frequency varied manually or by voltage control

When referring to crystal oscillators one usually considers a highly-stable fixed-frequency unit. Crystal oscillators can, however, be tuned. The amount of tuning is very small, typically less than 0.1%. Tuning of the oscillator tends to reduce the stability so that wide tuning negates the basic stability advantage of the crystal oscillator. The theoretical basis for tuning a crystal oscillator is quite simple when one remembers that the crystal represents a series LC circuit. This series circuit has very high Q; the motional inductance is large, the motional capacitance small, and circuit loss is low. Consider a typical case where the motional capacity is 0.01 pF. Here the addition of a 10-pF capacitor in series will reduce the equivalent series capacity by 0.1%. When the series capacitor is variable, it becomes possible to tune the crystal oscillator. Tuning can be either manual, or electronic by means of a voltage-controlled capacitance diode (i.e., a varicap). Such voltage-controlled oscillators are frequently referred to as VCXO's, Voltage-Controlled Crystal Oscillators.

In actual circuits, oscillator tuning is limited not only by stability considerations, but by crystal spurious responses. These spurious responses which, as previously discussed, can be quite close in frequency to the main response, make it difficult to tune crystal oscillators beyond fractional percentages.

MECHANICALLY TUNED OSCILLATORS

Most mechanically-tuned oscillators used in spectrum analyzers operate at microwave frequencies (several hundred MHz and up). We shall therefore concentrate on the special problems of these high-frequency units, excluding the many types of lower-frequency mechanically tuned oscillators.

With the exception of some relatively esoteric systems (e.g., avalanche diode), it is possible to classify microwave oscillators into two categories. One group, such as klystrons and magnetrons, is based on transit-time effects. We shall discuss these oscillators in a different section. The other group can be considered as working along the same principles as lower-frequency units, except that it is necessary to observe certain precautions. The following discussion pertains to this second group of oscillators.

transit time and stray reactance

Mechanically tuned microwave oscillators can be constructed using either tubes or transistors. At the present time tubes predominate in this area, so that our examples will be based on tubes. There are two basic limitations on the operation of electron-tube microwave oscillators. These are transit time and the effect of stray reactances. Changes in the transit time of the electrons through the interelectrode space may be thought of as changes in phase shift. This introduces difficulties in our ability to optimize oscillator feedback as the unit is tuned over wide frequency ranges. Although clever oscillator design can alleviate some of the problems, the basic problem belongs to the electron-tube designer. Thus, transit time is the main reason why electron tubes designed for microwave use have very small interelectrode spacings. Spacings of less than 0.005 inch are quite common. Effects due to stray inductance and capacitance are reduced by utilizing designs where these elements are incorporated into the tuning structure.

compensating L and C built into devices or tank

To illustrate the effect of stray elements consider the Hartley oscillator. At low frequencies we can neglect the effect of interelectrode and stray capacitance and stray lead inductance, resulting in the circuit of Fig. 6-3. At higher frequencies the stray elements cannot be neglected so that the circuit becomes that of Fig. 6-4. In particular, note that the stray cathode-lead inductance, L_k, elevates the cathode above ground potential, so we no longer have a Hartley oscillator. The various stray capacitances can be combined, and the stray cathode inductance can be combined with the external tuning inductance by running the oscillator in a grounded-plate configuration. The evolution of the equivalent circuit for the grounded-plate system is shown in Fig. 6-5.

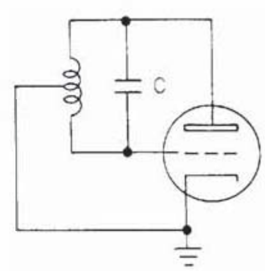


Fig. 6-3. Basic Hartley oscillator.

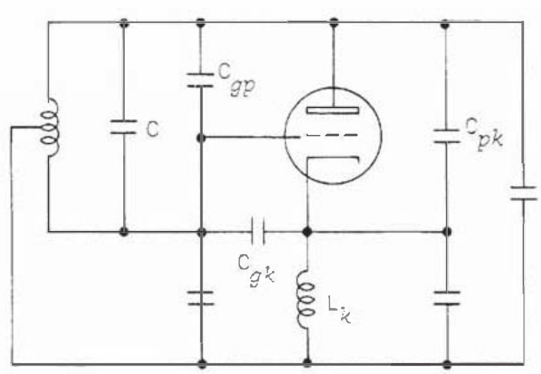


Fig. 6-4. Hartley circuit including stray elements.

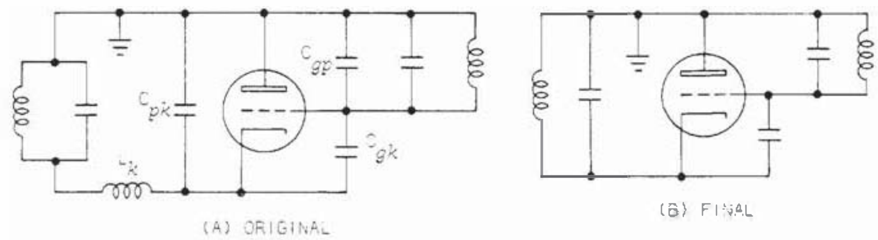


Fig. 6-5. Hartley oscillator grounded-plate configuration.

control stray reactance

Other configurations are the grounded-cathode oscillator shown in Fig. 6-6, and the grounded-grid oscillator shown in Fig. 6-7. The choice of circuit configuration depends on the physical and electrical properties of the electron tube. The final choice usually favors that configuration which permits easiest control of stray circuit elements. In any event all three configurations are quite similar. Each is composed of two tuned circuits and a feedback capacitor. Let us now consider methods for physically constructing such oscillators.

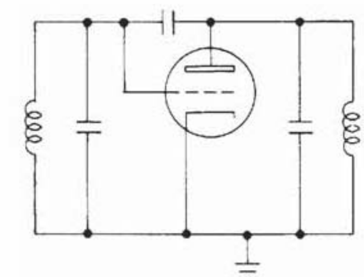


Fig. 6-6. Grounded-cathode oscillator.

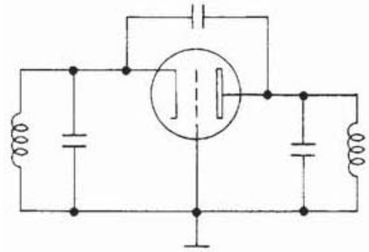


Fig. 6-7. Grounded-grid oscillator.

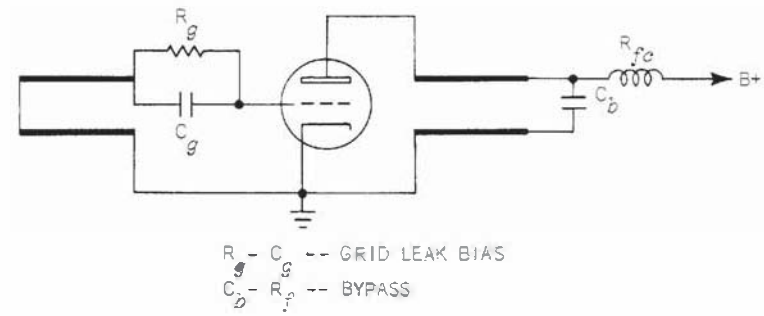


Fig. 6-8. Tuned-plate tuned-grid transmission-line oscillator.

$\omega = \frac{1}{\sqrt{LC}}$

quarter-wave tank

In each of these oscillators the operating frequency will be determined by the resonant frequency of the tank circuits: $\omega = \frac{1}{\sqrt{LC}}$. This leads to an LC product of about 2.5×10^{-20} for a 1-GHz oscillator. Obviously such a circuit cannot be constructed out of lumped elements. These types of oscillator are, therefore, constructed from transmission-line circuits. As discussed previously, a short-circuited transmission line one-quarter-wavelength ($\lambda/4$) long behaves as an antiresonant circuit. At the frequencies involved the use of short-circuited transmission lines is quite practical, since $\lambda/4$ at 1 GHz is about 3 inches. Our oscillator now acquires the basic circuit configuration shown in Fig. 6-8. Though it has been shown that this is a variation of a Hartley oscillator, we shall henceforth use the more common name -- Tuned-Plate Tuned-Grid, or TPTG for short.

TPTG oscillators can be constructed using various types of transmission line. At frequencies below 1 GHz it is quite common to use parallel lines for oscillator tuning. Such transmission lines are relatively simple and inexpensive. At the lower frequencies where the transmission lines become long it is not uncommon to coil the parallel line into a spiral or other configuration in order to save space. In many instances the oscillator is constructed out of a combination of lumped and distributed elements. The theory of operation of such combination oscillators is quite complicated, the design being more of an art than a science. One common technique is to add RF chokes to the filament lines. This helps to isolate the filament capacity from the

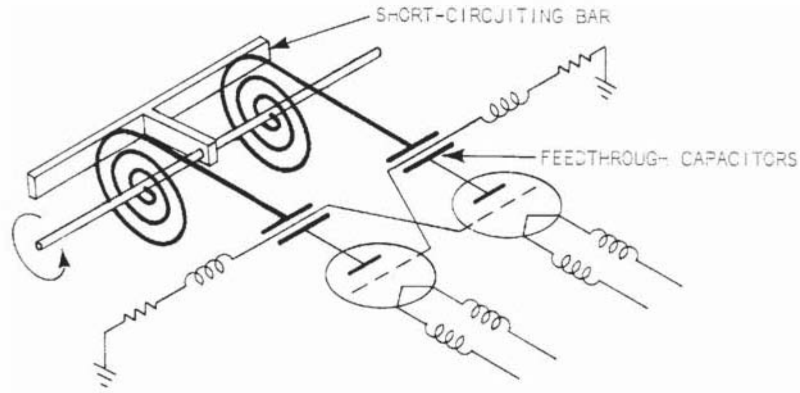


Fig. 6-9. Spiral-line oscillator.

active oscillator circuit. Fig. 6-9 shows a two-tube power oscillator with a spiral transmission line. The feedthrough capacitors provide the feedback. Oscillator tuning is by means of the shorting bar.

f > 1 GHz use coax

Above 1 GHz the transmission lines are almost invariably coaxial in nature. Coaxial lines have inherently high Q, closed coaxial lines exhibit low radiation and permit a design exhibiting good circuit isolation. One of the most popular coaxial-transmission-line oscillators is the grid-separation construction, so called because the grid line separates the plate line from the cathode line as shown in Fig. 6-10. The tuning structure consists of three cylindrical lines one inside the other. The middle cylinder serves as both the outer conductor of the cathode-grid coax line, and as the inner conductor of the plate-grid coax line. Shorting cylinders called *plungers* are moved between the coax lines to adjust their electrical length and so tune the frequency of the oscillator.

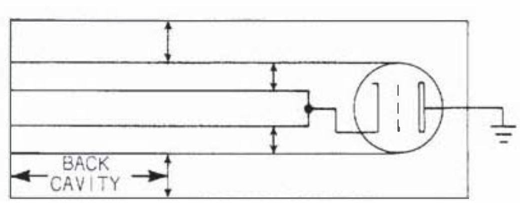


Fig. 6-10. Grid-separation coaxial-line oscillator.

designer has more control

The outside of the electron tube used in such an oscillator is of cylindrical construction, regardless of the electrode structure inside the tube. This permits a well-shielded structure, where the only connection between the two coax lines is via the tube interelectrode capacitance. Thus, with the normally low-capacitance tubes used it is necessary to provide intercoax feedback to make the oscillator work. This is a highly desirable situation as it permits greater freedom of design than would be the case if the feedback could not be controlled.

TPTG -- TP for frequency, TG for output

As in conventional units, TPTG-oscillator operating frequency is determined mostly by the plate-circuit tuning while the grid-circuit tuning controls power output and efficiency. The exact plunger setting for any frequency is normally determined experimentally. This is because the transmission line is not terminated by a perfect short. Rather we have the case of a transmission line somewhat less than $\lambda/4$ long resonating with the tube interelectrode capacity at one end, and the equivalent plunger capacity at the other end. Actually, as discussed in the section on transmission lines, the line does not have to be a quarter-wavelength long. The requirement for resonance is that the line be an odd number of quarter-wavelengths long. The quarter-wavelength condition is thus the first of many modes of operation. Often it is physically impossible to make the transmission line short enough for the desired frequency, but a shift to the $3\lambda/4$ mode solves the problem. Sometimes it is desirable to operate the two transmission lines at different modes, e.g., cathode-grid line at $1\lambda/4$ and plate-grid line at $3\lambda/4$. This is used as a cure for erratic operation caused by mode skipping. For example, at the same plunger setting the oscillator may run at the $\lambda/4$ mode at one time and $3\lambda/4$ at another time. Mode skipping is normally prevented by using a feedback structure that will couple mostly to the desired mode. When feedback optimization cannot be used, the two-mode tuning technique is used. The two-mode tuning technique is normally avoided as it introduces mechanical complexity, since the two lines have to be tuned at different rates.

multiple odd quarter-waves

plunger design critical

Proper design of the tuning plunger is of the greatest importance because not only must the plunger conform to certain electrical requirements, but it is also a mechanically moving part creating reliability problems. It is undesirable to use contacting plungers except in fixed-frequency or perhaps intermittent-operation oscillators, because tuning of the plunger causes wear and scarring of the metal surfaces. Changes in contact resistance cause tuning noise and metal particles get into the tuning mechanism causing erratic operation. To avoid these difficulties it has become standard practice to use noncontacting plungers.

capacitance plunger

choke plunger

coated surface plunger

isolation prevents coupling from back cavity

Noncontacting plungers are of many varieties. The simplest is an ordinary cylinder called a capacitance plunger. The most popular type is a variation of a choke plunger, called the S or Z plunger. In fact choke plungers are so prevalent that many people use the word *choke* as a synonym for plunger. Fig. 6-11 shows a cross sectional view of the capacity and S plungers. The plunger is isolated from the rest of the circuit either by a mechanical drive system that prevents the plunger from touching the walls, or as is more common, by coating the plunger with a low-loss nonconducting material such as Teflon or anodize. The reason for the prevalence of the S plungers, is that these plungers give the best isolation between the working oscillator and the back cavity behind the plunger. Good back-cavity isolation from the tuning (forward) cavity is important in order to prevent coupling between the two parts of the oscillator. Poor back-cavity isolation means that back-cavity resonances will affect oscillator tuning, an undesirable condition. Best isolation is obtained from a plunger which is a quarter-wavelength long at the frequency of interest. Plungers are therefore designed to be $\lambda/4$ long at the center of the tuning range.

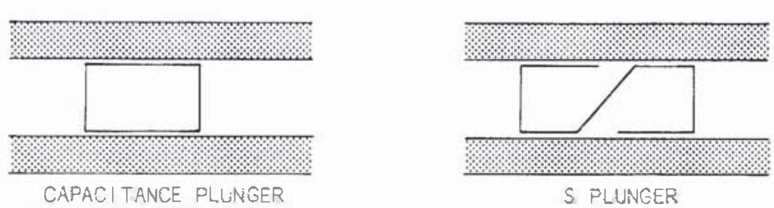


Fig. 6-11. Types of plungers.

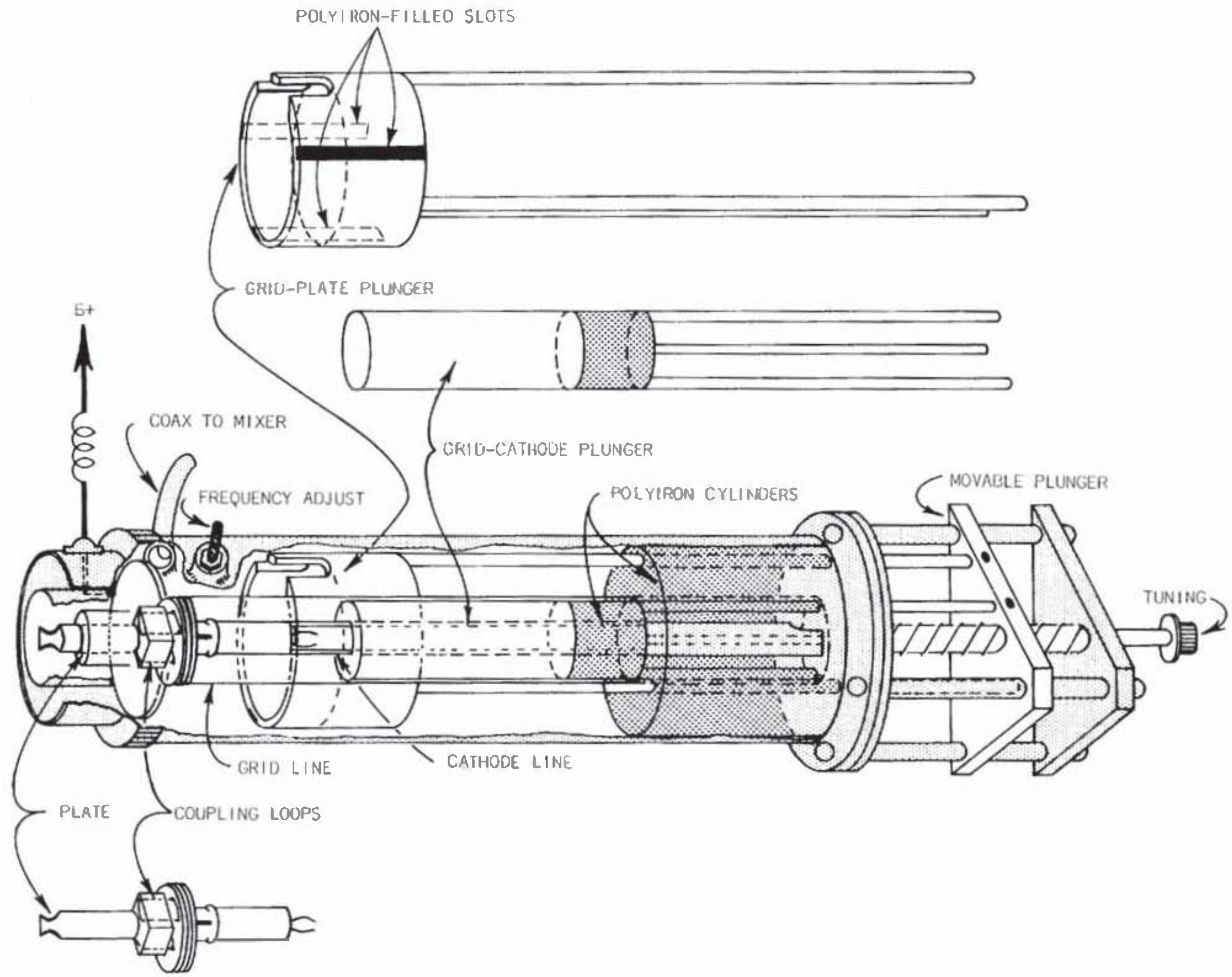
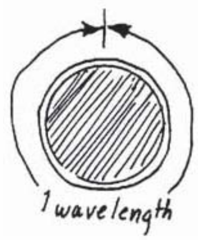


Fig. 6-12. Grid-separation transmission-line oscillator.

parasitic resonance



slot plunger with polyiron

One of the difficulties with noncontacting plungers is circumferential parasitic resonances. For example, there is a circumferential resonance at the frequency where the mean circumference of the outer gap is one-wavelength long. These resonances introduce dead spots where there is no oscillation and other tuning problems. These resonance effects are removed by slotting the plunger and loading the slots with lossy material. This both reduces the Q of the parasitic resonance, and moves the resonance down in frequency by virtue of the increased circumferential path. Lossy material, such as polyiron, is also used in the back cavity to reduce the effect of back-cavity resonances. The lossy material, in the form of a cylindrical plunger, is connected to the back of the metallic plunger and to the fixed back end of the oscillator. By this means both ends of the back cavity are terminated in absorptive lossy material, so as to prevent standing waves.

smooth silver surface

gap clearance

Two other considerations in the construction of transmission-line mechanically tuned oscillators are a good surface finish and silver plating to maintain a low-loss structure; and close tolerances, since the transmission-line-to-plunger gap must be on the order of a few thousandths of an inch. Also, eccentricities increase coupling to the circumferential resonances.

coaxial electron tube

The detailed construction of a TPTG grid-separation type of transmission-line oscillator is shown in Fig. 6-12. The electron tube is coaxial in structure. A nonmetallic holder supports the feedback wires which are looped through four holes drilled in the grid flange. The grid-plate plunger has three slots filled with polyiron lossy material for the suppression of circumferential resonances. The grid-cathode plunger has a piece of polyiron added on for the suppression of back-cavity resonances. In addition, back-cavity resonances are suppressed by fixed cylinders of polyiron at the back of the oscillator. The output power is coupled through a loop in the grid-plate line. An adjustable screw, which changes the circuit capacity and hence the frequency, is used for frequency adjustment.

TRANSIT-TIME OSCILLATORS

velocity modulate: bunching

For amplification or oscillation to take place it is necessary that the electron beam, originating at a cathode, be controlled or modulated at an RF rate. In an ordinary electron tube this is done by means of a grid, resulting in an amplitude modulation of the electron beam. In situations where transit-time effects preclude the use of amplitude modulation, *velocity modulation* is the general technique. Here the electrons in the beam are sorted into *bunches* by speeding up some electrons while retarding others. As the beam progresses from the point of generation (cathode) to the point of termination (anode), the bunching effect is reinforced at a particular frequency which is determined by the geometry of the tube and the electric and magnetic fields applied. The result is that more energy at bunching frequency can be extracted than has been put in. Hence we have an amplifier, or with appropriate feedback, an oscillator.

transit-time group

Transit-time oscillators are of many types. The best known are the *klystron, magnetron,* voltage-tuned magnetron (VTM), and backward-wave oscillator (BWO). Early spectrum analyzers made extensive use of the klystron. This is probably due to the early development of practical klystrons. Recently the VTM and BWO have entered the spectrum-analyzer field. These have the advantage of all-electronic tuning. However, the VTM has certain spurious-magnetic-resonance problems so that it is still not used much.

In the following discussion we shall give a brief description of the klystron, which is the best known and most instructive in basic principles, and the BWO which is being used in modern equipment.

A) The Reflex Klystron

reflex klystron single cavity

Early klystrons were of the multi-cavity type using all mechanical tuning. Spectrum analyzers use a somewhat later development known as the reflex klystron which uses only a single cavity or resonant circuit and is capable of limited electronic tuning. Fig. 6-13 shows the basic electrode structure of the reflex klystron. The electrons leave the cathode and are focused into a beam by the grid. The grid

also controls the size of the beam. The electrons are then accelerated by the high anode potential controlled by V_a. The electrons enter the resonator cavity at high velocity and pass through the cavity in a time interval which is short in comparison with the period of oscillation. Oscillations start as with any oscillator; when conditions are adequate, any small instability will start the self-sustaining process. When the klystron oscillates, RF fields appear within the cavity. Then the electrons in the beam, which pass through the cavity at an instant when the RF half-cycle accelerates them, will be speeded up; those that pass through in the other half cycle will be slowed down. The result is that the electrons enter the resonator in a relatively uniform beam but become bunched after passing through.

repeller -- dead end; go back to traffic light

The bunching is further increased by the action of the repeller as the electrons enter the *drift space* between the cavity and *repeller.* Here the electron beam is made to reverse direction by the retarding effect of the negative repeller. However, not all electrons will reverse direction at the same position in the drift space. Faster electrons will get closer to the repeller before reversing than will slower electrons. Thus, the faster electrons travel a greater distance than the slower electrons. The repeller voltage can be so adjusted that the faster electrons will be in phase with the slower ones as the electron beam reenters the cavity.

RF energy tapped

Here the RF energy of the highly bunched returning electron beam is extracted by some sort of coupling structure, shown as a loop in Fig. 6-13.

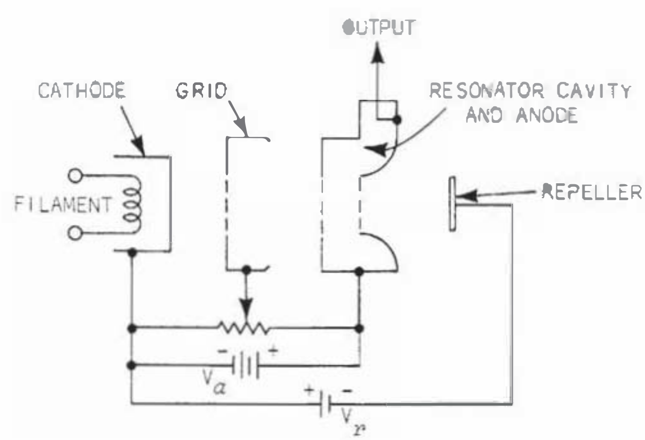


Fig. 6-13. Reflex-klystron structure.

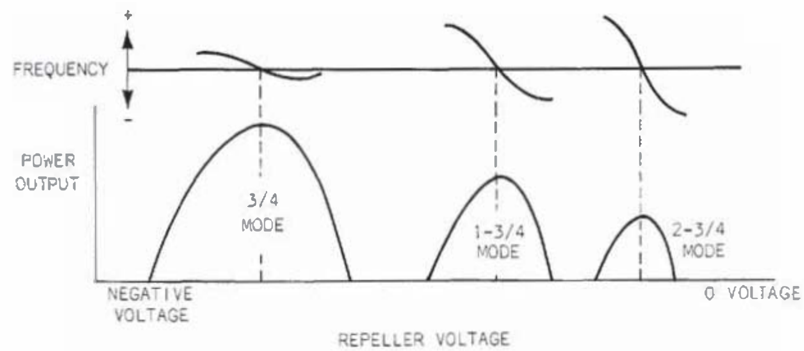


Fig. 6-14. Variations of frequency and power output as a function of repeller voltage.

repeller voltage compatible with cavity resonance

As previously indicated it is necessary that both the repeller voltage and the resonator resonant frequency be correct for self-sustaining oscillation to occur. Nevertheless, each of these can be tuned over narrow ranges without changing the other. This permits relatively narrow-range, up to about 30 MHz, electronic tuning of a klystron. Fig. 6-14 shows the power output and frequency variation as a function of repeller voltage for a reflex klystron. Note that the same frequency can be obtained at several different repeller voltage settings, where each repeller voltage corresponds to a different mode of oscillation. We have used the word *mode* to designate the electromagnetic field distribution in waveguides and transmission lines; to define resonant frequency-vs-physical dimensions relationship in waveguides and transmission lines; and now to differentiate between the various repeller voltage settings that result in the same output frequency in a klystron. Repeller modes arise because for optimum bunching reinforcement it is necessary that the electron beam be acted upon by the repeller for a time equal to (N - 1/4) cycles of oscillation. Thus we have a 3/4 mode, 1-3/4 mode, etc. corresponding to that repeller setting which causes the electron beam to spend a time equal to 3/4 of a cycle, 1-3/4 of a cycle, etc. in the drift space.

repeller modes

output coupling from resonator

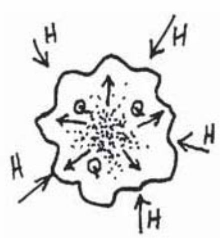
Physically the resonator is also the anode which contains the output coupling. As a result the anode is run at ground potential while the cathode is negative with respect to ground, and the repeller is negative with respect to the cathode. Typical voltages are about 300-to-1000 volts for the cathode and up to about 500 volts with respect to the cathode for the repeller. To operate a reflex klystron, it is therefore necessary to use a voltage-variable

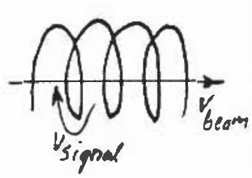
power supply in the kilovolt range, since the repeller voltage must be tuned. The resonator is either coax or waveguide depending on frequency. The construction is similar to that used in other types of oscillators. Thus, coaxial resonators use tuning plungers where the basic problems of type of plunger, circumferential resonances, etc., must be solved.

B) Backward-Wave Oscillator (BWO)

broadband tuning

A klystron is a narrowband device from the point of view of electronic tuning. In order to make a broadband electronically tunable device it is necessary that the beam-RF-field interaction be made broadband rather than like that of the narrowband resonator used in the klystron. In the BWO this takes the form of a helix on which propagates the RF field and a hollow electron beam inside the helix which interacts with the RF field all along the length of the helix. The electron beam is prevented from flying apart, due to the mutual repulsion of the electrons, by the focusing action of a magnetic field. This magnetic field can be either from an electromagnet or from permanent magnets spaced along the length of the helix. In order to produce a continuous interaction between the RF field and the electron beam it is necessary that the bunches on the beam remain in step, or in phase, with the RF field. Thus it is necessary to reduce the axial speed of the RF signal, which normally propagates at the speed of light, to the slower velocity of the electron beam. This is accomplished by having the RF signal propagate in a spiral, thus reducing its axial velocity.





frequency vs electron velocity vs cathode-to-helix voltage

The frequency of oscillation is determined by the velocity of the electrons only, since the electron beam-to-RF-signal interaction is broadband. Thus the frequency of oscillation is determined by the cathode-to-helix voltage which controls the electron velocity. BWO voltages run from about 200 volts to over a kilovolt. Frequency tuning ranges are from about 1.5-to-1 to 5-to-1. In construction the BWO consists of an electron gun or cathode, a helix, collector or anode, and a helix terminal or RF output. The reinforced RF wave travels in a direction opposite to that of the electron beam so that the helix terminal is at the side of the electron gun. The backward-traveling RF wave gives the oscillator its name.

RF wave goes backwards

Recently an attempt has been made to supplant the BWO by the voltage-tuned magnetron. The VTM has an inherent advantage since it has a linear voltage-frequency relationship while the BWO does not. However, there are still some basic problems with wide-tuning VTM's so that the BWO is presently the preferred component.

ELECTRICALLY TUNED OSCILLATORS

swept LO

electronically swept

One of the essential components of the superheterodyne signal-translating spectrum analyzer is a frequency varying, or swept local oscillator. The frequency sweeping could be performed by mechanical means, such as a motor-driven variable capacitor, however, reliability problems and restrictions in performance, such as low sweep rate, preclude the use of mechanical systems except in the most primitive designs. Therefore, we shall concern ourselves here only with electronically tuned or swept oscillators.

voltage-controlled resonant circuit

There are many types of electronically swept oscillators. One type, the BWO, was discussed in the section on transit-time oscillators. Almost every type of oscillator can, at least in theory, be converted to a voltage-controlled unit by the replacement of the usual resonant circuit with a voltage-controlled resonant circuit. At the present time the designer has available three electronically controlled frequency-selective elements. These are the voltage-controlled capacitance diode or varicap (also varactor), the current-controlled inductor, and the magnetic-field-controlled YIG. The most popular component, especially at frequencies below 1 GHz, is the varicap.

mod Pierce oscillator to VCXO

very small variable range

The varicap can be used to control the frequency of an oscillator either by taking the place of one of the frequency controlling capacitors or as an addition to an essentially completed oscillator. The addition idea is illustrated in Fig. 6-15. Here a Pierce oscillator, which is a variation of the Colpitts, is modified by the addition of the varicap. The effect of the varicap is to change the value of C2, this in turn changes the frequency of the oscillator. As discussed in the section on crystal oscillators, the frequency variation in this type of system is very small. Another example of a varicap-controlled oscillator is shown in Fig. 6-16. This oscillator is a variation of the familiar Colpitts

configuration and can be voltage tuned over large frequency ranges. Tuning of over 50% with respect to center frequency is quite common. One of the difficulties with wide-tuning-range varicap-controlled oscillators is the loss of a substantial portion of the varicap capacitance range due to RF bias. This occurs because most of the capacitance change of a varicap occurs over a small change of control voltage near zero volts. This aspect of varicaps is discussed in more detail in the section on frequency linearization. For the present it is sufficient to note that the RF, at oscillator frequency, appears at the varicap and determines the low-voltage limit to which the varicap can be biased by the control voltage. This in turn limits the capacitance change which limits the frequency tuning range. Many circuit variations, which at first glance may not make sense, are due primarily to an effort to keep the RF bias of the varicap to a minimum.

variable range limited by RF bias

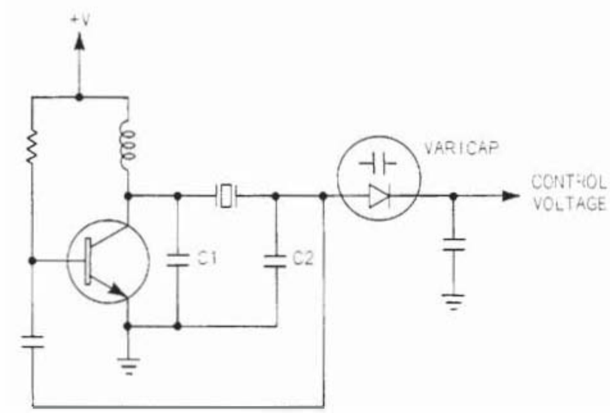


Fig. 6-15. Voltage-controlled crystal oscillator (VCXO).

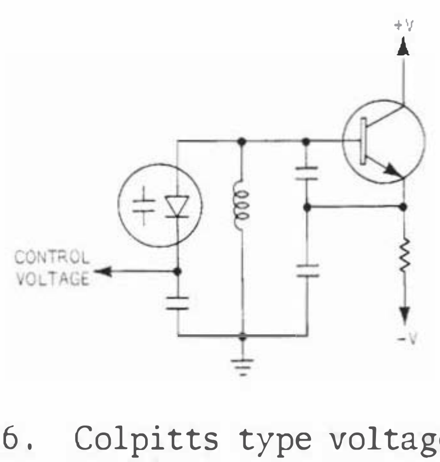


Fig. 6-16. Colpitts type voltage-controlled oscillator.

YIG resonator advantage

An electronically controlled oscillator technique with much promise for the future is that utilizing the properties of YIG resonators. This is particularly true at frequencies above 1 GHz where the properties of YIG resonators become quite attractive. The chief advantage of YIG is the inherent linear tuning, thus obviating the need for frequency-linearization circuits. At the time this was written transistorized YIG oscillators covering two-to-one frequency ranges up to 2 GHz have been announced in the literature. These units still have some problems, such as extreme sensitivity to load VSWR, and to spurious resonances. No doubt these problems will be solved as more work is done on this attractive component.

FREQUENCY LINEARIZATION TECHNIQUES

nonlinear voltage vs frequency

Most electronically controlled oscillators have a nonlinear control-voltage-versus-frequency relationship. This means that when a linear relationship is desired it is necessary to preshape the control-voltage tuning curve, or otherwise affect the oscillator circuit to make the voltage-frequency relationship linear.

Consider, for example, the use of a varicap-controlled oscillator. The capacitance-versus-control-voltage relationship is nonlinear, the greatest capacitance change occurring at the low end of the voltage range. A typical case is shown in Fig. 6-17. The problem is further complicated by the nonlinear frequency-versus-capacitance relationship, since the two are related by the resonance equation $f = \dfrac{1}{2\pi\sqrt{LC}}$. This relationship, normalized for L = 1 mH, is given in Fig. 6-18.

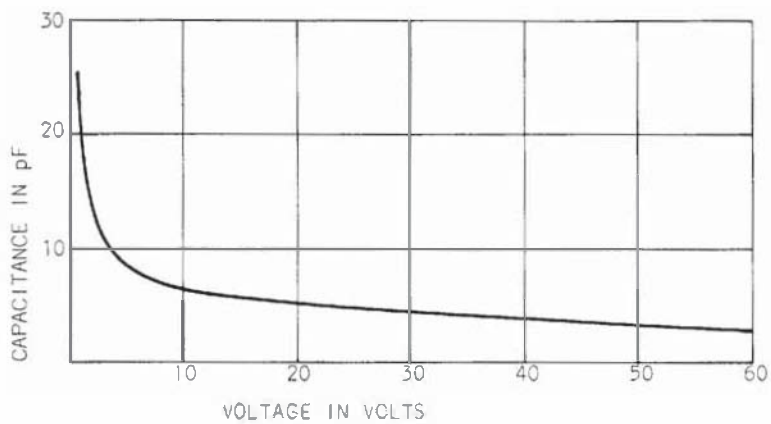


Fig. 6-17. Capacitance-vs-voltage curve for varicap.

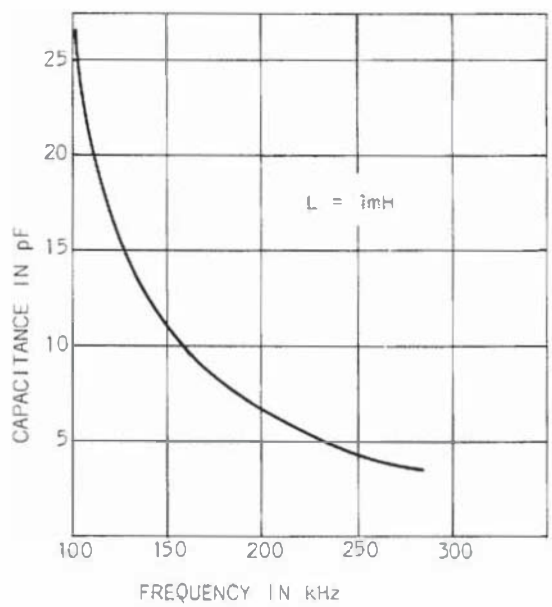


Fig. 6-18. Frequency-capacitance relationship for resonant circuit.

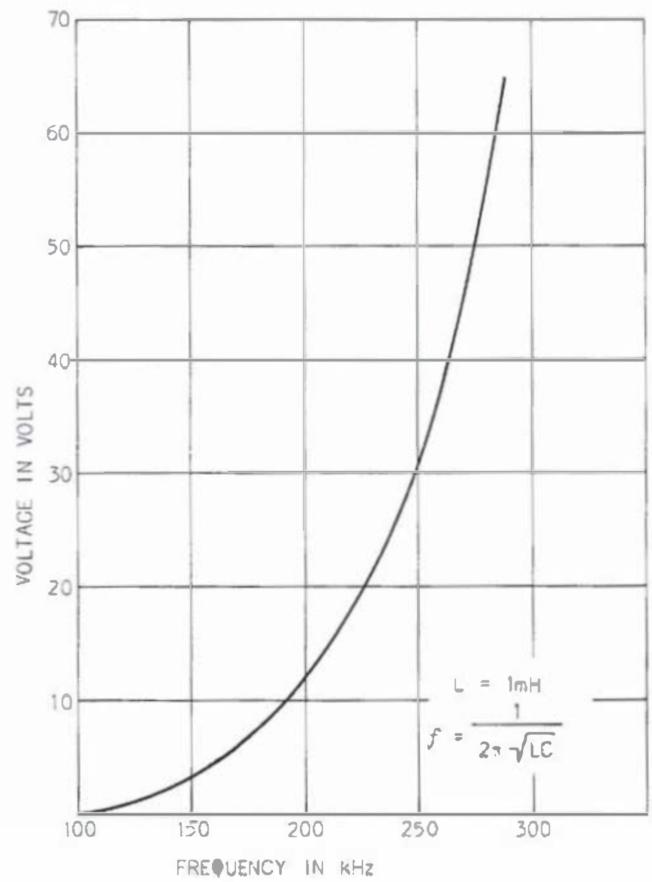


Fig. 6-19. Typical voltage-frequency relationship for a varicap-controlled oscillator.

Combining these two curves into Fig. 6-19 we observe that the voltage-frequency relationship for a varicap-controlled oscillator is not a linear one. The importance of Fig. 6-19 is much more than just an illustration of the nonlinear relationship between voltage and frequency. This curve also gives us the voltage-versus-time relationship when we wish to get a linear frequency-versus-time sweep. Thus, to convert a voltage sawtooth to a frequency sawtooth by means of a varicap-controlled oscillator we need to preshape the voltage sawtooth according to Fig. 6-19. This concept is illustrated in Fig. 6-20. Here we start with a linear *voltage* as a function of time and end with a linear *frequency* as a function of time. However, somewhere in the system we have had to produce a nonlinear *voltage*-versus-time relationship.

apply a nonlinear operation

The required voltage shaping can be obtained in two ways. One way is to use an open-loop shaper, such as illustrated in block form in Fig. 6-20. This technique assumes foreknowledge of the required shaping function. The other way is to use closed-loop feedback control of the shaping function. The open-loop method is the simpler and less expensive, while the closed-loop method provides more accurate control and best linearity.

open-loop shaper

closed-loop shaper

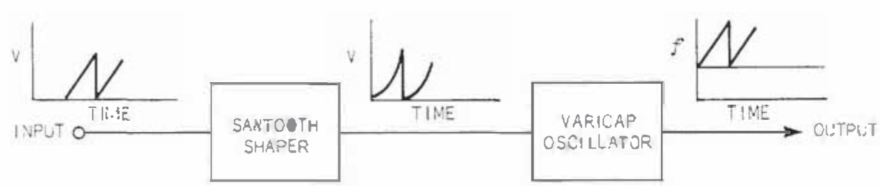


Fig. 6-20. Frequency sawtooth generator, block diagram.

Let us consider the open-loop method first. We assume foreknowledge of the input and required output voltage-time functions. This is illustrated in Fig. 6-21, where the known input and the computed or experimentally determined output are plotted on the same graph. We can approximate the shape of any portion of the output curve simply by reducing the slope of the input curve. All that a reduction of slope really means is a reduction in voltage and this can be accomplished by a passive divider network. What we need is a set of resistive voltage dividers which are switched in or out when the appropriate voltage is reached. As an illustration consider a system where, as the voltage ramp starts from zero, a ten-to-one divider is switched in so that the output is one tenth as large as the input. After a fixed time interval, when the input voltage has reached a certain predetermined value, the voltage divider is changed into a five-to-one, then a ten-to-three, a ten-to-four, etc. Fig. 6-22 shows what happens in such a system. The upper curve was constructed by using the above division ratios. The addition of voltage offset to eliminate the vertical steps leads to the smoother lower curve. The ideal curve can be approached as closely as desired simply by increasing the number of breakpoints. Of course, an infinite number of breakpoints does not make for a very practical arrangement, so real systems usually have between three and ten breakpoints.

segment slope $= \frac{V_{in}}{t_{1,2,3,\ldots n}}$ for n segments

more steps -- smoother curve

Prebiased diodes make excellent voltage-sensitive switches, so the divider network usually consists of a diode-resistor matrix. Depending on the desired degree of control of the breakpoint voltage and output slope, a divider may contain up to five resistors and two diodes. A ten-breakpoint network could, therefore, contain up to fifty resistors and twenty diodes just in the voltage-divider matrix. In real circuits the divider network is usually operated in conjunction with an amplifier. This permits slope control without regard to the maximum slope of the input waveform; DC restoration problems associated with the biased divider networks are avoided, and by connecting the divider network as a feedback loop around the amplifier we convert the difficult problem of constructing a passive network whose loss decreases with increasing input to the simpler problem of a network where the loss *increases* with increasing input. The final system, though based on simple theoretical principles, can therefore get quite complex.

diodes switch the steps

limitations avoided by using amplifier with feedback

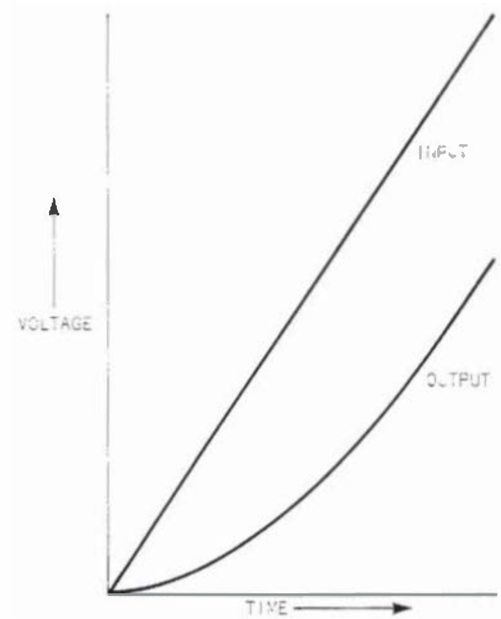


Fig. 6-21. Voltage-time relationship for a sawtooth-shaping circuit.

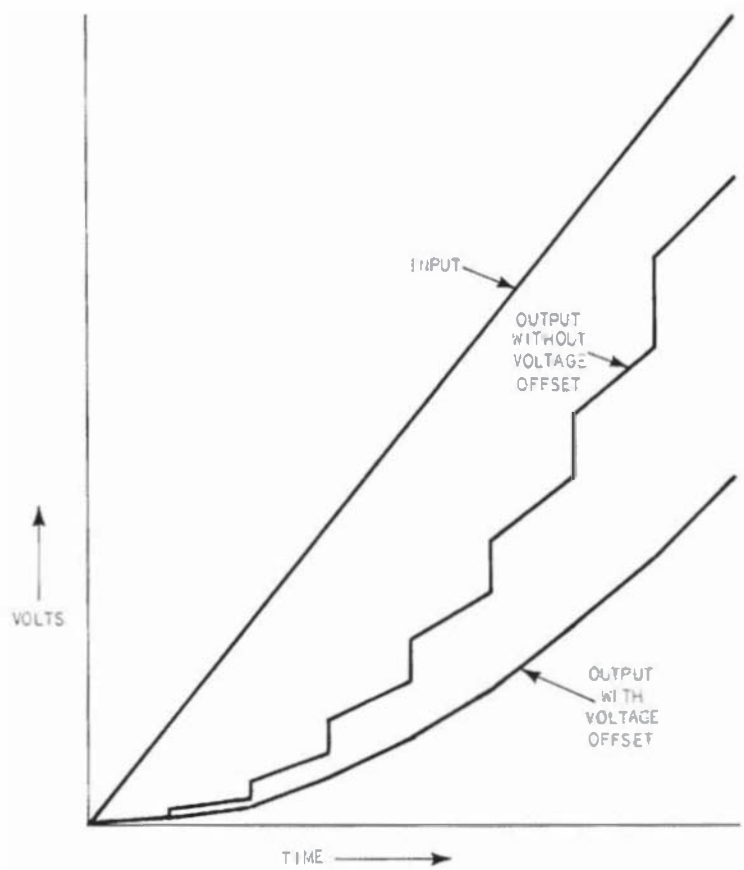


Fig. 6-22. Piecewise linear-shaping outputs.

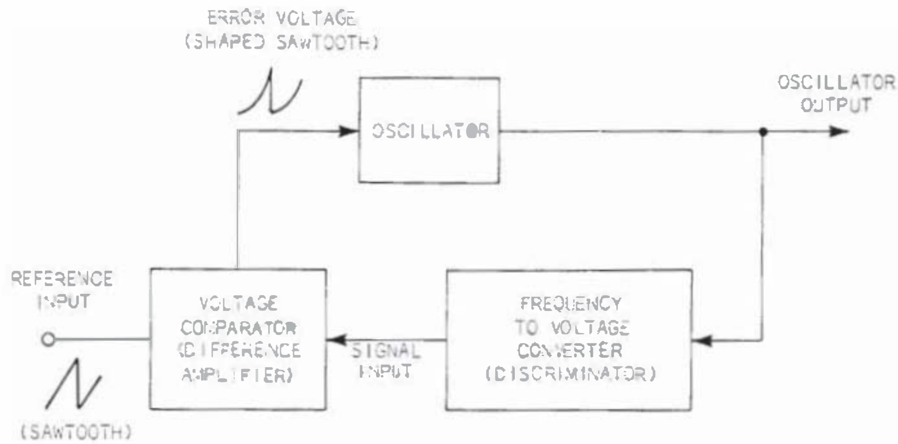


Fig. 6-23. Block diagram of feedback-controlled sweeping oscillator.

closed-loop feedback

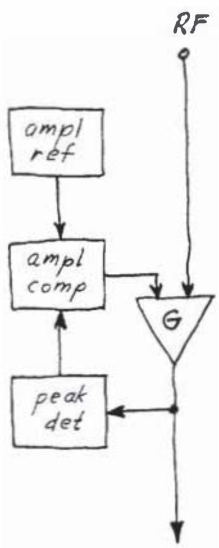
An alternate oscillator-frequency-linearization technique uses a closed-loop feedback-controlled system. The oscillator output is compared to a reference signal and the resulting error signal is used to control or correct the oscillator. Fig. 6-23 is a block diagram of the basic system.

As discussed in the section on closed-loop-controlled systems, given a sufficiently high-gain comparator we can neglect the minute difference between the reference input and the signal input needed to generate the error output. For all intents and purposes the controlled output has the same characteristics as the reference input, and the oscillator output follows the voltage-frequency characteristics of the discriminator. It is, therefore, important to consider the problems associated with the frequency-to-voltage conversion.

discriminator senses input voltage

maintain input V constant

The discriminator circuit was discussed elsewhere, so we will treat the discriminator as a black box that converts a frequency-varying input into an analog voltage-varying output. However, there is one problem: The discriminator is not only frequency sensitive, it is also voltage sensitive. This means that variations in input voltage appear also as variations in output voltage. This in turn feeds false information to the voltage comparator, resulting ultimately in undesired frequency variations in the oscillator output. It is, therefore, essential that the voltage of the frequency-varying input signal to the discriminator be maintained at a constant level. This is accomplished in a similar manner to the frequency control, namely by a feedback-controlled system.



The amplitude-control loop consists of a peak detector, which converts the RF frequency-varying signal to a DC voltage having the same amplitude-vs-time variations as the RF signal; a voltage comparator amplifier, which compares the detected voltage with a fixed amplitude reference; and an output amplifier, whose gain is controlled by the comparator error-signal output. As with other closed-loop feedback-controlled systems it serves no useful purpose to consider the detailed waveforms appearing in different parts of the system. Our only concern is that the system have sufficient gain so that effectively the output is controlled to have the same characteristics (in this case constant amplitude) as the reference signal. Of course, this is all predicated on the assumption that the RF-output amplitude variations are faithfully reproduced by the peak detector. However, this is not always the case when the oscillator output contains harmonics. Let us, therefore, consider the effect of signal harmonics on peak detectors.

A good peak detector has no frequency discrimination within broad limits. The output voltage will, therefore, be proportional not to the peak of the fundamental of the input waveform, but to the peak of the composite waveform including the fundamental and all the harmonics. Thus, a constant-amplitude fundamental is made to look unflat by the addition of varying amounts of harmonics. This is illustrated by the photographs in Fig. 6-24. These photographs illustrate what happens to the waveform when varying levels of 2nd and 3rd harmonic are added to a constant-amplitude fundamental sinewave.

wideband peak detector

responds to composite peaks

We note that the addition of even harmonics distorts the sinewave but has no effect on the overall amplitude. Odd harmonics, on the other hand, not only distort the waveform but also change the peak-to-peak amplitude. Ordinarily it is of little consequence for mixer operation if the local-oscillator signal contains a small amount of third harmonic. But, in the case of a closed-loop feedback-controlled oscillator, this can cause substantial frequency nonlinearity. Observe that as the odd-harmonic content of the waveform changes, the output of the peak detector will follow, and the amplitude-control loop will provide appropriate correction signals to maintain a constant amplitude. However, if the fundamental content of the oscillator output were constant, then the effect of the corrective action of the amplitude loop would be to vary the amplitude of the fundamental. The discriminator in the frequency-control loop, being a frequency sensitive device, responds to the fundamental so variations in fundamental amplitude at the input will result in amplitude variations at the discriminator output. The discriminator-output voltage controls the oscillator frequency-versus-time function, so that odd harmonic content in the oscillator output causes considerable frequency-versus-time nonlinearity. The even harmonics are not as objectionable since the peak-to-peak amplitude is not affected. But these also cause problems because the DC level of the waveform is shifted from zero by the vertical asymmetry in the waveform, as illustrated. Furthermore some discriminators, particularly those using transmission lines, are subject to the multiple resonances of transmission lines and are adversely affected by harmonics. This is why a low-pass or bandpass filter is frequently added to the control loop. The function of the filter is to reduce the harmonic content of the oscillator output.

even and odd harmonic effects

discriminator responds to fundamental

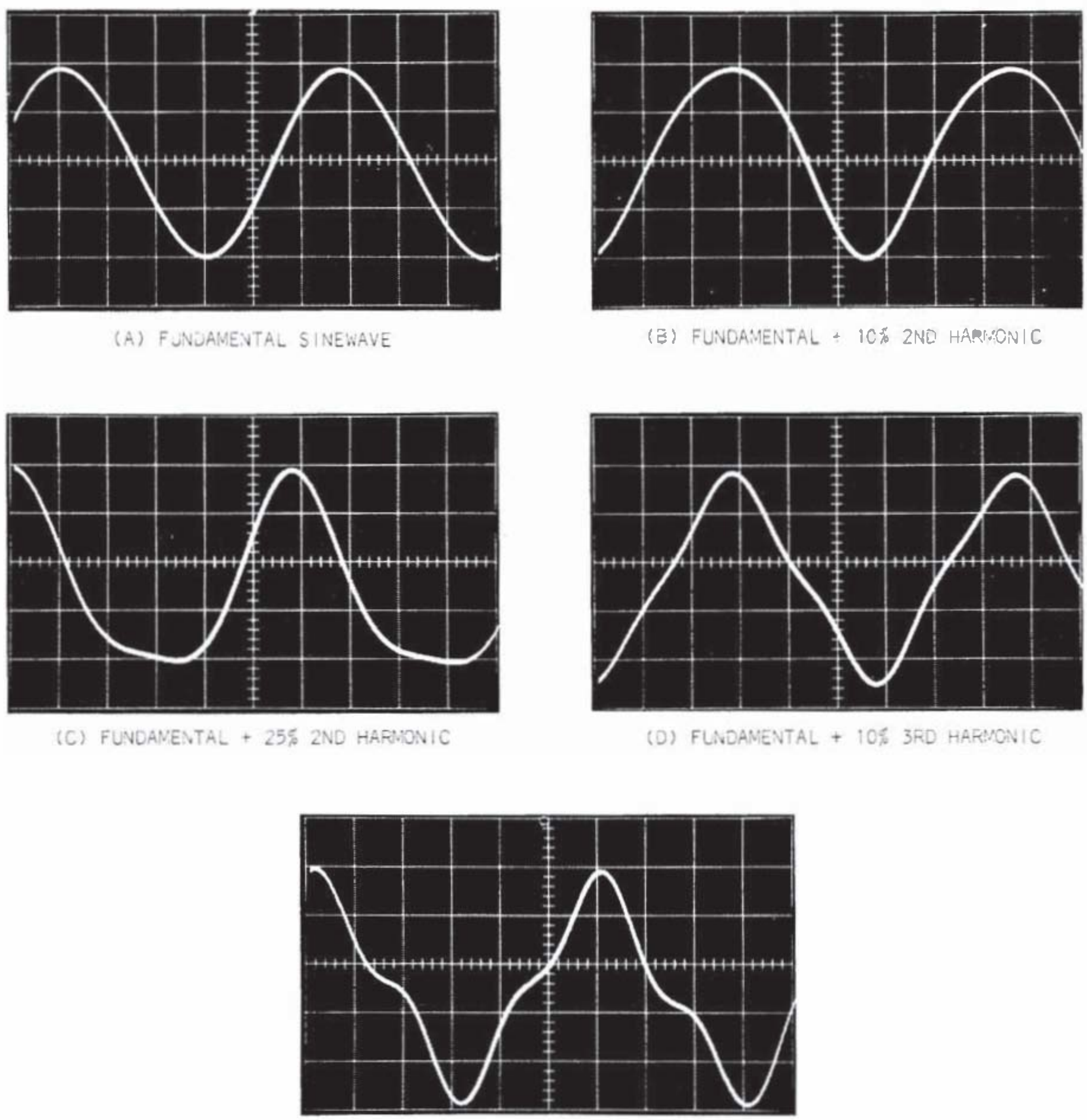


Fig. 6-24. Effect of harmonic content on waveform.

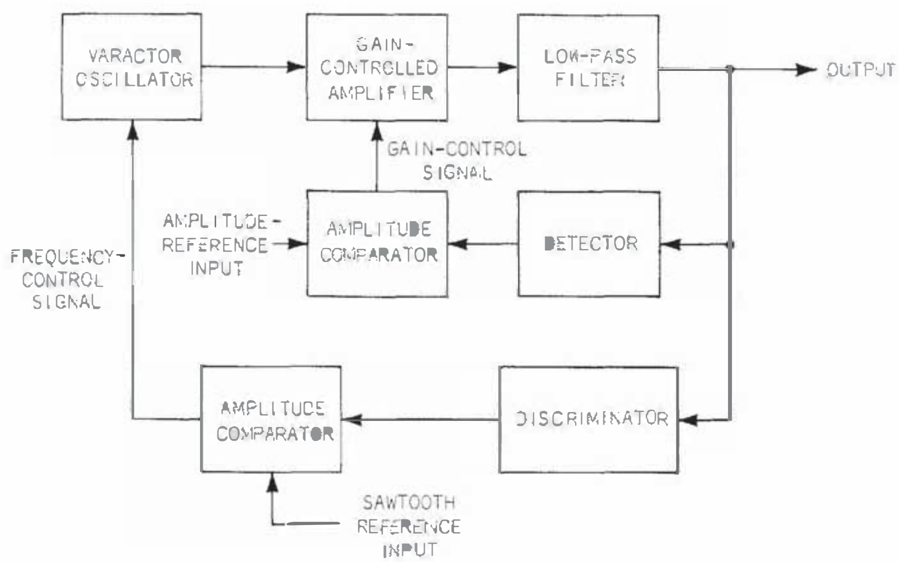


Fig. 6-25. Complete block diagram of feedback-controlled swept oscillator.

A block diagram of the complete frequency-control system including the amplitude loop is shown in Fig. 6-25.

PHASELOCK OSCILLATOR STABILIZATION

feedback-control signal determined by environment

Besides improvements in linearity and accuracy in frequency adjustment, another important advantage that the feedback-linearization system has over the open-loop shaper technique is improvement in oscillator stability. In the open-loop system nothing can be done about unpredictable oscillator-frequency deviations caused by power-supply noise, vibration, temperature changes, and other uncontrollable factors. This is because the system operation is based on foreknowledge of the required control signal. In the feedback system, however, the control signal is determined by the conditions prevailing at the time of operation. Thus, if the oscillator frequency should for some reason deviate from the required value, the feedback system will generate an error voltage of such magnitude as to restore the frequency to its original magnitude.

system stability limited by reference, discriminator and comparator

The discriminator-controlled feedback system is usually capable of only limited improvement in oscillator-frequency stability. This is because the ultimate stability of the oscillator cannot be made better than the ultimate stability of the reference system. The reference system is not only affected by the reference signal, but by the discriminator and comparator as well. In the world of stable oscillators where short-term stabilities of one part in 10^8 are not uncommon, the discriminator-controlled system is several orders of magnitude below what can be achieved. The greatest degree of stability is achieved by a feedback-controlled system where the reference is a stable frequency source, the comparator consisting of a phase detector. This type of system is known as a phaselock loop or phaselock system.

phaselock provides superior stability

The basic phaselock system was described in the section on closed-loop control circuits. Fig. 6-26 is reproduced from there. The reference strobe consists of a stable train of narrow pulses generated by a crystal oscillator working in conjunction with a narrow-pulse generator, such as described in chapter two. The sampling gate is one of the several types of phase detectors and samplers as described in the section on phase detectors. The difference amplifier needs no discussion and the phase-control signal is a stable variable DC voltage which sets the oscillator frequency.

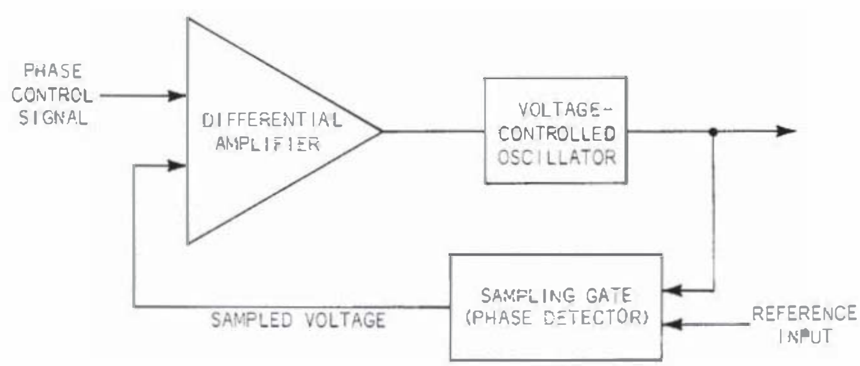


Fig. 6-26. Phaselock-controlled oscillator system.

The voltage-controlled oscillator is not necessarily the type of voltage-controlled oscillator described in the section on electronic tuning. Two types of oscillators were described there; the very narrow-frequency-range crystal-controlled unit and wide-tuning transistor oscillators. Actually all oscillators can be voltage controlled to some extent. The output frequency of a tuned-plate tuned-grid oscillator is determined, among other things, by the stray and tube interelectrode capacity. A varactor lightly coupled to the plate-grid transmission line can, therefore, be used for voltage-tuned frequency control.

step-tuned phaselock

Phaselock oscillator systems are of two types, the step-tuned oscillator and the continuous-tuned oscillator. The step-tuned system is the simpler of the two; see Fig. 6-26. The controlled oscillator is tuned, by mechanical or other means, to a frequency which is harmonically related to the reference input. At that time the phase-detector output is a particular predetermined value, usually zero. Any deviation of the oscillator frequency from this preset value will generate a phase-detector output which will in turn restore the oscillator to the frequency which is an exact multiple of the reference frequency. Once the oscillator is phaselocked the frequency is fixed. In this system, the oscillator frequency cannot be tuned while maintaining phaselock.

continuously tuned phaselock

The continuously tuning phaselock system is comprised of the same basic elements as the step-tuned system. The difference between the two is that in the continuously tuned system the oscillator can be voltage controlled over a wide range of frequencies, while the reference frequency is swept. Imagine, for example, a wide-tuning 1-to-2-GHz oscillator such as a BWO or transistor-varicap unit. This oscillator is to be phaselocked to a reference which tunes from 1-to-2 MHz. As long as the microwave oscillator tunes in synchronism with the reference source, the 1000:1 (1 GHz is 1000 times 1 MHz) relationship between the oscillators will be maintained, and the 1-to-2-GHz oscillator will continue to be phaselocked during tuning. In practice the system is more complicated since wide-tuning stable-reference oscillators are very difficult to make. Actual systems use a relatively narrow-tuning stable

oscillator as a reference, and continuous tuning of the high-frequency oscillator is maintained by using several reference harmonics. For example, we wish to phaselock a 1-to-2-GHz oscillator to a 1-MHz reference source that can be tuned through a range of only 1 kHz: At 1 GHz the oscillator is locked to the 1000^{th} harmonic of 1 MHz. At 1 MHz above 1 GHz (1.001 GHz) the oscillator is locked to the 1000^{th} harmonic of 1.001 MHz. The reference-oscillator frequency immediately returns to 1 MHz and the microwave oscillator locks to the one-thousand-and-first harmonic of 1 MHz. This process continues until the microwave oscillator at 2 GHz locks to the one-thousand-nine-hundred-and-ninety-ninth harmonic of 1.0005+ MHz.

5 control loops needed

The continuously tuning, or sweeping, phaselock system is much more complicated than the step-tuned system. The sweeping system needs a sweeping reference source which can be accurately tracked to the high-frequency oscillator. This usually means closed-loop-discriminator frequency control for both oscillators. The result is a total of five control loops: Two amplitude-control loops, two discriminator-control loops and one phaselock loop. There are also problems associated with the need to prevent the oscillator from locking to the wrong reference harmonic. In spite of the problems, the stability improvement of phaselock over other systems make the sweeping phaselock system very attractive.

acquisition

hold-in

pull-in

The remainder of this section is devoted to a discussion of two important phaselock properties, *acquisition* and *hold-in*. So far it was assumed that the high-frequency oscillator is somehow tuned to a harmonic frequency of the reference source and phaselock occurs. This is not always easy since the *pull-in* frequency range can be quite narrow. The need for getting the high-frequency oscillator within the pull-in range of the phaselock loop, a process called acquisition, accounts for many otherwise puzzling features of phaselock systems. One technique is to provide a manually tuned fine-frequency control. This control in conjunction with a beat-frequency phaselock indicator permits manual tuning into the pull-in range. The so-called phase-control signal in Fig. 6-26 can be thought of as an electronic fine-tuning control which helps in phaselock acquisition.

A more sophisticated technique is narrowband sweeping or frequency *dither*. Here the phase-control signal is swept electronically, thus getting the oscillator frequency within the pull-in range without intervention of the operator. The swept technique is particularly useful in situations where phaselock may be broken during a measurement or other spectrum analyzer use. Here phaselock is quickly reestablished without operator intervention. Of course, in the sweeping system it is important to restrict the acquisition sweep width to less than the hold-in range, otherwise phaselock could not be maintained. Finally, an auxiliary discriminator-controlled loop can be used to bring the oscillator frequency within the pull-in range.

hold-in range versus loop time-constants

A sufficiently strong disturbance will cause a phaselocked oscillator to lose lock. The capability of a phaselocked oscillator to maintain lock is defined by the hold-in range. Hold-in range depends on the loop time-constants, so a slow disturbance may not break phaselock while a fast disturbance of equal magnitude may. For example, time-constant considerations put an upper limit to the sweep rate in the swept acquisition system. Other circuit characteristics affecting hold-in range are: Loop gain, the greater the gain the wider the hold-in range; loop bandwidth, which is connected with the time-constant question mentioned earlier; oscillator-control range, pertaining to the frequency range over which the oscillator can be voltage controlled; and circuit saturation, pertaining to the maximum control voltage that can be handled. Saturation sets the ultimate limit on hold-in, since it would, for example, serve no purpose to increase the loop gain to the point where the circuit saturates on input noise.

REFERENCES; see page 171.

Oscillators -- B-11, B-12, B-13
Magnetrons -- B-2
Klystrons and Triodes -- B-3
Transmission Line -- B-12
Crystal -- C-2 pp66-70, C-7 pp117-118
Phaselock -- C-2

7

RF ATTENUATORS

The purpose of an attenuator* is to reduce the power delivered from a source to a load when the attenuator is inserted between them. Thus, the major characteristic of an attenuator is the *insertion loss*. Insertion loss can be defined in several different ways, the definition requiring that the attenuator be inserted between a matched source and load is the most prevalent. The requirement of a matched source and load make the insertion-loss specification a characteristic of only the attenuator. Thus, we define insertion loss "L", or *attenuation* as it is sometimes called, as

$$L = 10 \log_{10} \frac{P_{in}}{P_{out}} \text{ dB}$$

with source and load matched. It is important to recognize that the above definition does not necessarily require that the attenuator dissipate power in order to produce attenuation. For example, a dissipationless metal obstruction properly placed in a transmission line will reduce the amount of power delivered from the generator to the load by reflecting some of the power back to the generator. This type of attenuation is called *reflective attenuation* and is based on different principles than *dissipative attenuation*. A prime example of a reflective attenuator is a waveguide beyond cutoff, while the best known example of a dissipative attenuator is a simple resistive-divider network. In the following sections we shall consider the characteristics of some of the types of attenuators that may be used with spectrum analyzers.

not all attenuators dissipate power

* The word *pad* is sometimes used as a synonym for attenuator. Usually though, a pad is considered an uncalibrated power-reducing device, whereas an attenuator is calibrated.

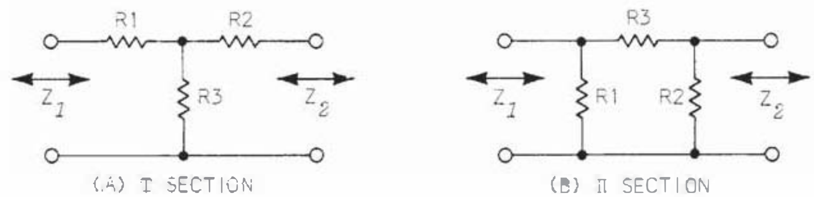


Fig. 7-1. Resistive attenuators.

RESISTIVE-DIVIDER ATTENUATORS

The simplest type of attenuator is that based on the dissipation of a resistor or resistors connected in series and/or shunt with the transmission path. Though any arbitrary arrangement of resistors along the transmission path will reduce the power delivered to the load, at least three resistors are necessary for a controlled-impedance match between the attenuator and the load and source. Two basic three-resistor sections are used. These are the T section and the Π section as shown in Fig. 7-1. As shown, the attenuators are unsymmetrical; that is, R1 ≠ R2 and Z1 ≠ Z2. In most applications the source and load operate at the same impedance so that Z1 = Z2 and hence R1 = R2. Further, the impedance of the transmission system is usually 50 Ω, so most designs are based on a 50-Ω system.

Consider as an example a 50-Ω 10-dB T attenuator. There are three specified parameters; Z1 = 50 Ω, Z2 = 50 Ω, and $\frac{P_{in}}{P_{out}} = 10$ or $\frac{V_{in}}{V_{out}} = \sqrt{10} = 3.16$. This leads to three equations with three unknowns, so the attenuator resistors are uniquely specified. The result is an attenuator where R1 ≃ 26 Ω, R2 = R1 ≃ 26 Ω, and R3 ≃ 35 Ω.

The attenuator arrangement is shown in Fig. 7-2. The attenuator consisting of two 26-Ω and one 35-Ω resistors is connected to a 50-Ω load and driven by a signal source whose internal resistance is 50 Ω and open-circuit voltage is V_o. The input resistance is 26 Ω in series with the combination of 35 Ω in parallel with 50 Ω and 26 Ω in series. This leads to $Z1 = 26 + \frac{(35)(26 + 50)}{35 + 26 + 50} \simeq 26 + 24 = 50$ Ω.

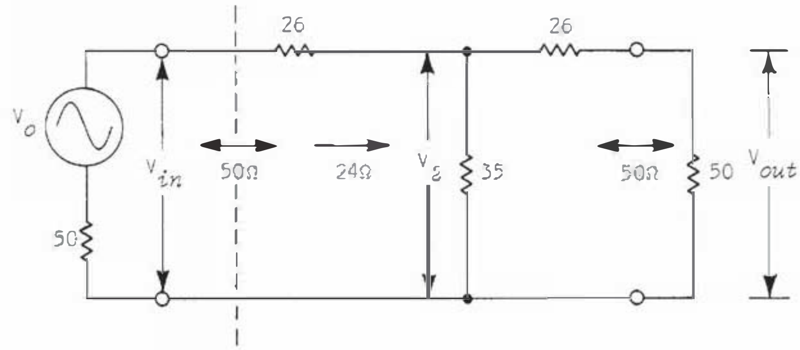


Fig. 7-2. 50-Ω 10-dB T attenuator.

A computation for the output impedance leads to the same result. Note that even though the load is not part of the attenuator proper, it is considered as part of the circuit for the purpose of impedance calculation. Similarly the output impedance of the generator is taken into account when computing the attenuator-output impedance.

$\frac{P_{out}}{P_{in}}$

Looking from the generator terminals we see a 50-Ω load connected to a 50-Ω generator. Therefore, by simple voltage division, the input voltage, V_{in}, is one half of the signal-generator open-circuit voltage, V_o. This loss in voltage is not counted as part of the attenuator effect, since the same loss will occur when the generator is connected directly to the load with the attenuator removed. Looked at in another way, it is desired to compute the input-to-output ratio of power or voltage. Starting with V_{in} the voltage divides between 26 Ω and 24 Ω resulting in $V2 = \frac{24}{50} V_{in}$. This voltage V2 appears across the 50-Ω load in series with 26 Ω so that $V_{out} = \frac{50}{76} V2$. The result is $V_{out} = \frac{24}{50} \cdot \frac{50}{76} V_{in} = 0.316\ V_{in}$, or 10-dB down. Note that both load and source are connected to the same impedance in both the presence and absence of the attenuator. The only effect of the attenuator has been to reduce the power delivered to the load by a factor of 10 with respect to the power delivered to the load when the load and source are connected directly together. One-tenth of the power delivered by the source goes to the load and nine-tenths is dissipated by the resistors of the attenuator.

Note that slightly more than half of the input power is absorbed in the first 26-Ω resistor in a 10-dB T attenuator. This resistor must be capable of dissipating this power, otherwise the attenuator will be damaged.

At relatively low frequencies, where lumped circuit elements are used, the attenuator can be constructed in almost any configuration. All that one must be careful about is accuracy, power dissipation and stability of the resistors. At higher frequencies, however, where distributed circuits are used, the construction technique becomes important. Attenuators having good performance characteristics up to 1 GHz can be constructed with ordinary quarter-watt and half-watt carbon resistors. The three resistors forming the T or Π configuration are connected in a more or less coaxial configuration where the series resistors form part of the center conductor. This type of construction cannot be used much above 1 GHz because of the influence on insertion loss and input impedance of stray capacity and inductance, and the discontinuities in the transmission structure due to the large size compared to a wavelength of the resistors.

physical relationships critical over 1 GHz

At frequencies in the gigahertz region, attenuators are constructed from microwave elements commonly called *rod* and *disc* resistors. These are distributed resistors formed by depositing a conductive film on a nonconductive substrate such as pyrex glass. The names rod and disc are based on physical appearance, as shown in Fig. 7-3. The rod diameter is controlled to form the center conductor of a transmission line having a specific characteristic impedance. Likewise the diameter of the disc is adjusted to mate with the outer conductor of the desired transmission line. This type of structure works well to about 8 GHz, and has been used successfully to about 12.4 GHz.

resistive film on rods and discs

A technique that works quite well from DC to beyond 12.4 GHz is that based on a single conductive card. It is still a resistive-divider attenuator, though hardly recognizable as such. Here the distributed nature of the resistors is one step beyond the rod-disc concept, since both the shunt and series resistors have been combined into one distributed resistor.

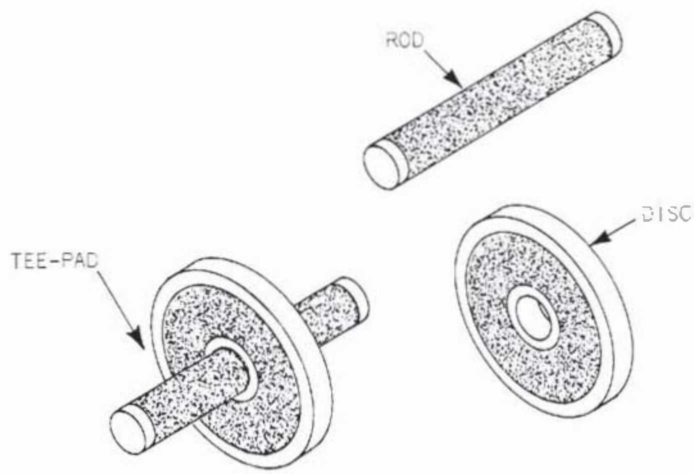
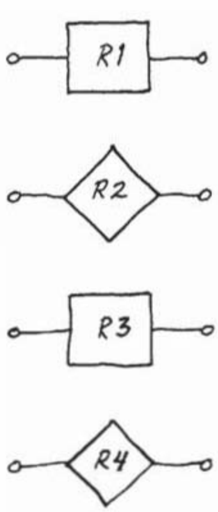


Fig. 7-3. Rod and disc resistors.

resistance per square

Before proceeding further, it is helpful to discuss the concept of *surface resistivity* or as it is often called *ohms per square*. Consider a long thin conductor of length ℓ and cross-sectional area A. We measure the end to end resistance of this wire and find it to be R. Further we know that R is proportional to the length ℓ, if ℓ is doubled R also doubles. This follows directly from the addition of resistors in series. Likewise it follows from the analogy of resistors in parallel that R is inversely proportional to A. Adding a proportionality constant ρ (rho) we have an equation $R = \frac{\rho \ell}{A}$, where ρ is the *resistivity* of the material from which the conductor is made. Dimensionally, resistivity has the units of resistance-length. One common unit of ρ is ohm-cm. Imagine now that the conductor is in the form of a thin sheet rather than a wire. This sheet has width "w", thickness "t", length "ℓ", and the resistivity of the material "ρ." Then the resistance measured across the length of the sheet is $R = \frac{\rho \ell}{A} = \frac{\rho \ell}{wt}$, since the cross-sectional area is thickness times width. If we have a square, where $\ell = w$, the resistance R is only dependent on the material thickness and resistivity. Regardless of the size of the square the resistance R remains the same, and this is called ohms per square.



$R_1 = R_2 = R_3 = R_4$ for same ρ and t

Normally when discussing surface resistivity it is necessary to include the frequency of operation. This is because at high frequencies the current

RF goes only skin deep

flows near the surface of the conductor; the depth of major current flow is called the *skin depth**. Thus, if the skin-depth δ is less than the material-thickness t it is the skin depth and not the material thickness that should be used in computing the surface resistivity. Since skin depth varies inversely as the square root of frequency, it is usually necessary to specify frequency when discussing surface resistivity, e.g., a four-fold increase in frequency cuts the skin depth in two. However, the conductive films on attenuator cards are made extremely thin so that the ohms per square remains constant over the operating-frequency range.

A conductive card attenuator is constructed by connecting a diamond-shaped card between input and output center conductors and the outer shell of a coaxial transmission line as shown in Fig. 7-4.

The resistance card has the appropriate ohms-per-square resistance and the proper size and shape to provide the desired attenuation while maintaining a good impedance match. The amount of attenuation is dependent on the length of the card so that quite large attenuation values can be obtained in single card attenuators. Sixty-dB attenuators of this type are available on a routine basis, while in the rod-disc construction anything above 30 dB is considered special. The problem of making large and small attenuators out of discrete components, such as rods and discs, is that of controlling stray inductance and capacitance. A 30-dB 50-Ω T attenuator has two 46.9-Ω series resistors and a 3.2-Ω shunt resistor. A very small amount of stray inductance in series with the 3.2-Ω resistor will destroy the properties of the attenuator.

---

* Skin depth, denoted by δ (delta), is the depth where the current has decreased to 1/e the value at the surface. It is also the depth where, if the current were uniformly distributed, the surface resistivity would be the same as for the actual case where the current has an exponential distribution.

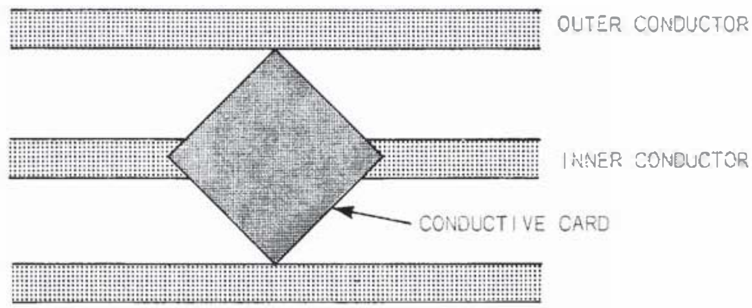


Fig. 7-4. Conductive-card coaxial attenuator.

WAVEGUIDE BEYOND-CUTOFF ATTENUATOR

As shown in Fig. 1-8 and reproduced here as Fig. 7-5, waveguides have a *cutoff* frequency, passing energy above the cutoff frequency and attenuating all inputs below the cutoff frequency. One way of looking at it is that a waveguide is a high-pass filter, attenuating all the inputs at frequencies below the cutoff frequency. The high-attenuation-frequency region below cutoff is the basis for a versatile attenuator variously known as a *piston attenuator* or *cutoff attenuator*. Cutoff attenuators can be constructed in any kind of waveguide, but circular waveguide

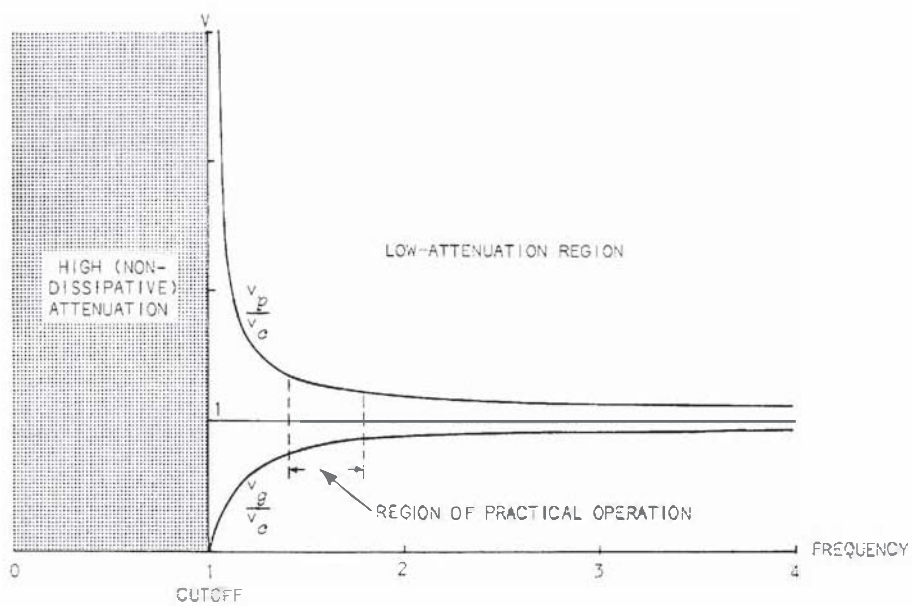


Fig. 7-5. Normalized phase velocity and group velocity vs normalized frequency.

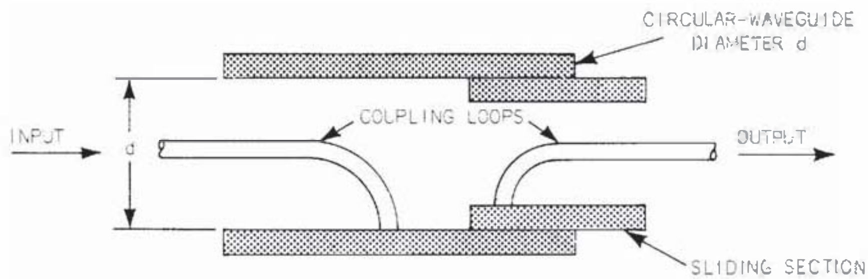


Fig. 7-6. $TE_{1,1}$-mode cutoff attenuator.

accurate variable attenuator

permits easier mechanical construction. Though fixed insertion-loss attenuators can be constructed, the main advantage of the cutoff technique is that it leads to highly accurate variable attenuators. The basic method of construction is illustrated in Fig. 7-6. Power is coupled into the guide by means of a fixed loop which excites the $TE_{1,1}$ mode of propagation. The power propagates down the guide with a high degree of attenuation, and the small percentage that gets to the other end is coupled out with a second loop. The amount of attenuation is determined by the distance between the loops, which is controlled by sliding the output loop with respect to the input loop. The coupling structure does not necessarily have to be a loop, though loop coupling is the most popular method. For example, an alternate method utilizes discs. The major difference between the disc and loop method is that the disc structure excites the $TM_{0,1}$ cutoff mode as opposed to the $TE_{1,1}$ cutoff mode for the loop.

It has been shown that for frequencies well below cutoff the insertion loss of a waveguide beyond cutoff is given by:

$$L = \frac{54.58}{\lambda_c} \text{ dB per unit length.}$$

λ_c = cutoff wavelength

For the $TE_{1,1}$ mode, $\lambda_c = \frac{3.41}{2}$ d, where d = diameter of circular guide, so for the $TE_{1,1}$ mode the insertion loss is about 32 dB per guide diameter. For example, if the coupling loops are separated by three waveguide diameters the attenuation is 96 dB.

In the above formula for attenuation to be valid, there are two basic requirements: One, previously mentioned, is that the operating frequency be well below cutoff; the other requirement is that only one

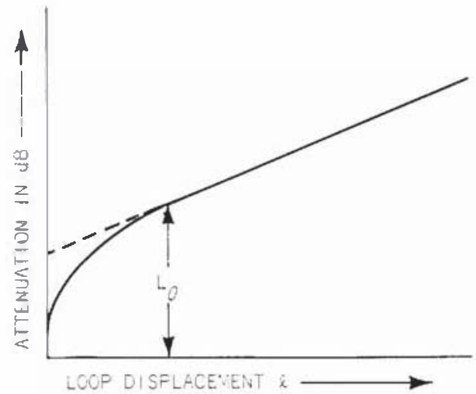


Fig. 7-7. Representative calibration curve for a cutoff attenuator.

mode be excited in the waveguide. When the loops are close together, however, the mode pattern is disturbed so that at small loop separations the attenuation-versus-distance relationship becomes nonlinear as shown in Fig. 7-7. Thus, in most applications, the cutoff attenuator is permitted to have a fixed insertion loss denoted as L_0 in Fig. 7-7. Cutoff attenuators are therefore not used in situations where high insertion loss, as high as 30 dB, cannot be tolerated. The great advantage of these attenuators is that, at attenuations greater than L_0, the attenuation-versus-displacement characteristic can be very accurately predicted, the accuracy of the attenuator being almost completely determined by the accuracy of the mechanical structure. Therefore, one major use of cutoff attenuators is as a standard against which other attenuators are checked.

high insertion loss

Cutoff attenuators are nondissipative, the input power being reflected back toward the generator. As a result these attenuators can handle much greater power levels than resistive attenuators which are based on power dissipation. Power dissipation for resistive attenuators is measured in watts, whereas that for cutoff attenuators is hundreds of watts. One consequence of the reflective as opposed to dissipative nature of cutoff attenuators is that the input impedance is reactive. This can cause application difficulty since some circuits, for example oscillators, will not function properly unless terminated in a specific impedance. This problem is aggravated by the use of long transmission lines

reactive input impedance

between the attenuator and other circuits. A transmission line acts as a frequency-sensitive impedance transformer, thus the input impedance is not only reactive but varies with frequency as well. Solutions include the use of very short transmission lines or coupling oscillator circuits directly to the attenuator loop, and the use of external resistive pads. The use of external resistive pads, however, cuts the power-handling capability of the attenuator.

MISCELLANEOUS ATTENUATOR CIRCUITS

lossy lines as attenuators

One type of dissipative coaxial attenuator is a length of lossy transmission line. Such transmission line, described in chapter one, can have considerable loss at higher frequencies. A better controlled version of the same thing uses a nonconductive center conductor covered with a resistive metalized film. The attenuation is controlled by the ohms per square of the conductive film. The drawbacks of this type of attenuator are changes in attenuation with frequency, and the fact that the attenuator does not have a DC return to ground, which is necessary for certain circuits such as mixers.

waveguide attenuators

Besides the reflective cutoff attenuator there are also lossy attenuators in waveguide. Two major types are the *flap* attenuator and the *rotary-vane* attenuator.

The flap attenuator is illustrated in Fig. 7-8. A lossy resistance card is inserted through a slot in the center of the broad wall of a rectangular

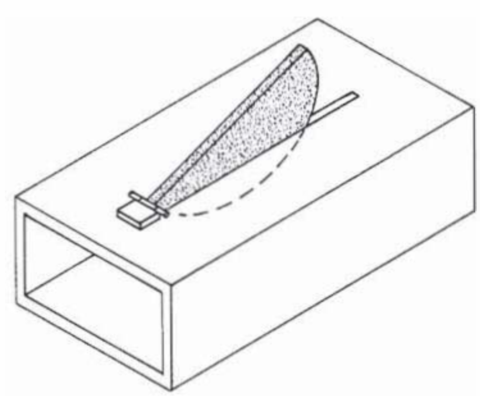


Fig. 7-8. Flap attenuator.

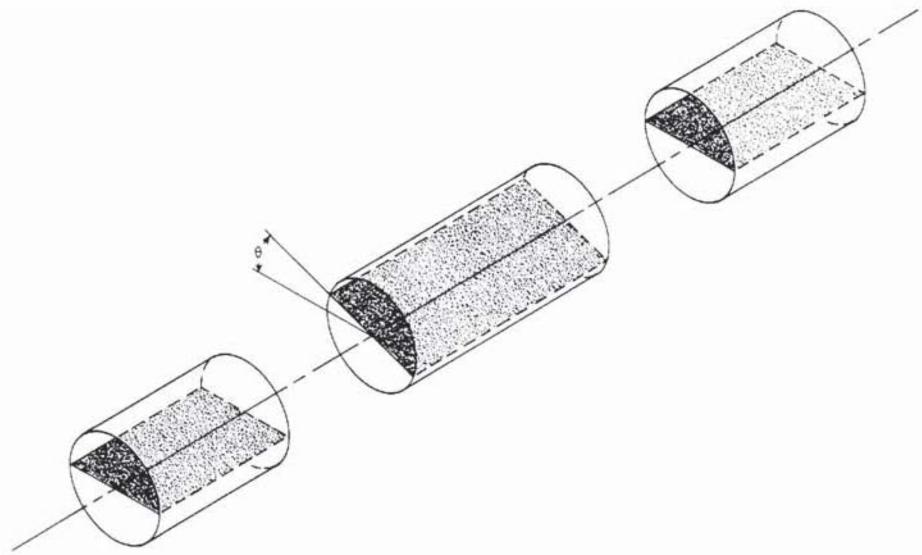


Fig. 7-9. Rotary vane attenuator.

waveguide. The deeper the insertion the more the attenuation. Though the flap can be shaped to give a more or less linear attenuation curve, this attenuator suffers from frequency sensitivity. On the other hand, the rotary-vane attenuator is inherently flat with frequency and the attenuation is accurately predictable according to a mathematical formula. This attenuator consists of three sections of waveguide, each having a lossy-resistance card down the center as shown in Fig. 7-9. The two end sections are fixed with respect to each other with the lossy cards in the same plane. The center section, however, is rotatable. The end sections have rectangular-to-round waveguide transitions so that the direction of electric field is perpendicular to the lossy card. As the center section is rotated, the central card changes direction with respect to the electric field, thus changing the amount of insertion loss. Negligible loss occurs when all the cards are in the same plane and perpendicular to the electric field. Maximum loss occurs when the center card is parallel to the electric field. The attenuation is, thus, determined by the angular position of the central card with respect to the other two cards. The ratio of input to output signal varies as $\cos^2 \theta$, where θ is the angle between the vanes as shown in Fig. 7-9; in decibels the insertion loss is:

$$L = 20 \log \frac{1}{\cos^2 \theta} = 20 \log \sec^2 \theta$$

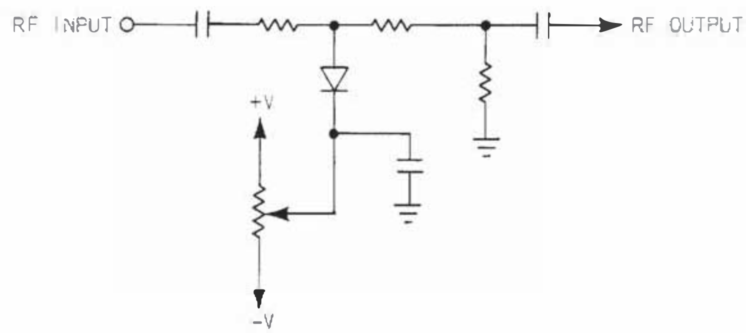


Fig. 7-10. Simple diode attenuator.

diodes as attenuators

Besides lossy center conductors in coax and lossy cards in waveguide, dissipative attenuators are also constructed from diodes. The advantage of diode attenuators is that diode resistance and hence attenuation can be controlled electronically. Almost any diode can be used as a rudimentary attenuator as illustrated in Fig. 7-10. When the diode is back biased it represents a high impedance and the attenuation is low. As the diode becomes forward biased the shunt impedance at RF frequencies is decreased and the attenuation increases. For wide-range attenuators several diodes are utilized. A diode that is particularly suited to modulator and attenuator applications is the PIN diode. These diodes have appreciable storage time so they do not rectify at frequencies above about 100 MHz. The result is a voltage-controlled resistor that can be used in all of the resistive-attenuator circuits discussed previously. PIN-diode modulators and attenuators are available commercially from several hundred megahertz to the upper gigahertz region.

high-Z attenuators

Some spectrum analyzers, especially those operating at low frequency, may use high-impedance attenuators. These are based on the same principles and use the same circuits as attenuators used in oscilloscopes. The high-impedance attenuators are straightforward voltage dividers, the only complication being the need to compensate for stray input and output capacity. Thus, the attenuator is usually constructed from fixed resistors and variable capacitors which are adjusted to compensate the attenuator.

Several types of attenuators having adjustable insertion loss have been discussed so far. All of these have been of the continuously variable variety. Sometimes these attenuators are not suitable for the application, so that a different kind of variable insertion-loss attenuator has to be utilized. These are constructed by combining a set of fixed-loss attenuators with appropriate switching. Such *stepped* or *switched* attenuators are available in differing configurations depending on the type of attenuators and switches utilized. Switches include diodes, toggle switches, rotary switches, and other mechanical configurations. Some arrangements stack attenuators in cascade, thus achieving the maximum attenuation steps with the fewest attenuators, while other arrangements utilize one attenuator at a time. Switching can be either electrical or mechanical, and there are many readout techniques for telling the operator what his total attenuation is.

REFERENCES; see page 171.

Attenuators -- B-13 Chapter 9
Microwave -- B-11 Chapters 11 and 12, B-12 pp1016-1023

BIBLIOGRAPHY

The purpose of this bibliography is to provide sources of material for more background study of the topics presented. No attempt has been made to give credit to the original developers of the subject matter presented. This bibliography is not an index to all the books and articles published on the subjects covered in this volume.

The bibliography contains three types of references:

A. *Background References* covering most of the prerequisite material necessary for a full understanding of spectrum-analyzer circuits. This is particularly true in regard to microwave circuits, where some readers may be deficient.

B. *General References* contain detailed material on most of the subjects covered.

C. *Specialized References* have been included to cover the areas not treated adequately in the general references.

Each reference is listed at the end of each chapter to which it is applicable, and in this bibliography where a short summary is included.

A. *Background References*

1. Camps & Markum. *Microwave Primer*. Bobbs-Merrill Co., 1965.
An elementary treatment of microwaves, presented in a question and answer format. The book stresses transmission-line and waveguide concepts, and includes a discussion of various microwave components and measurement techniques.

2. Wheeler, G.T. *Introduction to Microwaves*. Prentice-Hall, Inc., 1963.
A basic textbook in microwave theory and techniques. No mathematics beyond algebra is used. Topics include Transmission Lines, Microwave Measurements, Waveguides, Coaxial Lines, Methods of Matching, Tees and Couplers, Microwave Components, Resonant Cavities and Filters, Mixers and Detectors, Switching Techniques, Antennas and Microwave Tubes.

3. Lance, A.L. *Introduction to Microwave Theory and Measurements.* McGraw-Hill, 1964.
A somewhat more theoretical text than number 2 above. Topics include Elementary Field Theory, Transmission Lines, Use of the Smith Chart, Waveguides, Waveguide Components, Frequency Measuring Devices, Ferrite Devices, Active Microwave Devices, Attenuation Measurements, Measurement of Microwave Power, and Microwave Measurement and Calibration Techniques.

B. *General References*

The basic general reference is the 28-volume M.I.T. Radiation Laboratory series. McGraw-Hill, 1947-1953. Boston Technical Publishers, 1964. These books, though written some twenty years ago, have lost none of their timeliness or relevance to today's problems. Some of the topics pertinent to this book are:

1. *Pulse Generators.* Volume 5.

2. *Microwave Magnetrons.* Volume 6.

3. *Klystrons and Microwave Triodes.* Volume 7.

4. *Principles of Microwave Circuits.* Volume 8.
Includes a chapter on Resonant Cavities.

5. *Microwave Transmission Circuits.* Volume 9.
Includes an extensive discussion on Microwave Filters.

6. *Waveguide Handbook.* Volume 10.

7. *Technique of Microwave Measurements.* Volume 11.
Many important topics including applications of Spectrum Analyzers. Contains a detailed discussion of Microwave Attenuators and Directional Couplers.

8. *Crystal Rectifiers.* Volume 15.

9. *Microwave Mixers.* Volume 16.
With Volume 15, presents all the basic theory of the Crystal-Diode Mixer and associated topics.

10. *Vacuum-Tube Amplifiers.* Volume 18.
Modern spectrum analyzers seldom use Vacuum-Tube Amplifiers, but the basic relationships of Tuned Circuits, Sensitivity, and Noise Figure are applicable to all types of amplifiers.

11. *Microwave Receivers*. Volume 23.
Includes Mixers, Local Oscillators, IF Amplifiers, Video Amplifiers, Detectors, and other important topics.

The second general reference source is the two-volume Harvard series. McGraw-Hill, 1947.

12. *Very High-Frequency Techniques*.
These books contain much design information. In particular Volume II is the prime reference on Transmission-Line Oscillators.

The last of the general references is:

13. *Reference Data for Radio Engineers*. 4th ed., ITT, 1956.
The more than 1000 pages contain data on a tremendous range of subjects. This reference is particularly appropriate in the areas of Resistive Attenuators and basic Filter Design.

C. *Specialized References*

1. Del Toro and Parker, *Principles of Control-Systems Engineering*. McGraw-Hill, 1960.
All detailed treatments of Feedback Control Systems use complex mathematics, and this book is no exception. About one-hundred pages of this 686-page book is devoted to a development of the required mathematics. Detailed study of Feedback Control Systems is not recommended for those who have difficulty with mathematics.

2. Gardner, F.M. *Phaselock Techniques*. John Wiley & Sons, 1966.
Less mathematical. Includes applications of Feedback Control Systems. This book presents an up-to-date treatment of Phaselock Techniques and allied subjects such as Phase Detectors, Crystal Oscillators, Discriminators, and others. The book contains an extensive bibliography of over 100 references grouped according to subject matter.

3. Matick, R.E. "Transmission-Line Pulse Transformers Theory and Applications," *Proc IEEE*. January 1968.
This paper should be studied by all who wish to have a better understanding of Transmission-Line Transformers.

4. Lewis and Wells. *Millimicrosecond Pulse Techniques*. 2nd ed., Pergamon Press, 1959.
This book presents a thorough grounding in the basic principles involved in Narrow-Pulse Generation.

5. Hamilton and Hall. "Shunt-Mode Harmonic Generation Using Step-Recovery Diodes," *Microwave Journal*. April 1967.
A helpful paper in understanding how the Step-Recovery Diode may be used to generate narrow pulses.

6. Thiess, G.M. "Theory and Design of Tuneable YIG Filters," *Microwaves*. September 1964.

7. Pappenfus, Bruene, and Schoenike. *Single-Sideband Principles and Circuits*. McGraw-Hill, 1964.
Among other things this book includes a useful discussion of Intermodulation, Crystal Filters, and Balanced Mixers.

8. Matthaei, Young, and Jones. *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*. McGraw-Hill, 1964.
This eleven-hundred-page book is a major reference in the design, construction, and theory of operation of Filters, Directional Couplers, and YIG-Tuned Filters.

9. Rheinfelder, W.A. *Design of Low-Noise Transistor Input Circuits*. Hayden Book Co., 1964.

10. Thornton, et al. *Multistage Transistor Circuits*. John Wiley & Sons, 1965.
The book presents a detailed discussion of Transistor-Amplifier Design.

11. Woroncow and Croney. "A True IF Logarithmic Amplifier Using Twin-Gain Stages," *The Radio Electronic Engineer*. September 1966.
Presents a detailed discussion on Logarithmic Amplifiers.

12. Engelson, M. "Performance of Harmonic Diode Mixers," *Microwaves*. August 1967.

13. Wilds, R.B. "Microwave Two-Phase Converters for Imageless Receivers," *Microwave Journal*. September 1961.
Basic theory of Imageless Mixers.

14. Zverev, A.I. *Handbook of Filter Synthesis*. John Wiley & Sons, 1967.
Presents up to date information on filter synthesis including crystal filters and elliptical filters.

INDEX